DOCUMENT RESUME

ED 190 647 TM 800 444

AUTHOR	Bossone, Richard M., Ed.
TITLE	Proceedings. The National Conference on Testing: Uniting Testing and Teaching (3rd, Washington, D.C., March 13-14, 1980).
INSTITUTION	City Univ. of New York, N.Y. Center for Advanced Study in Education.
REPORT NO	Case-03-80
PUB DATE	Mar 80
NOTE	134p.: For previous proceedings see ED 141 404, ED 152 585, and ED 166 245.
AVAILABLE FROM	Graduate School of CUNY, Dean Richard M. Bossone, 33 W. 42nd St., New York, NY 10036 ($5.00)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Achievement Tests: Basic Skills: College Entrance Examinations: Diagnostic Tests: *Educational Assessment: *Educational Testing: Elementary Secondary Education: Federal Programs: Higher Education: *Information Utilization: *Instruction: Minimum Competency Testing: *Teacher Attitudes: Testing Problems: Writing Skills
IDENTIFIERS	National Institute of Education: Test Publishers: Truth in Testing Legislation

ABSTRACT

Following opening remarks by Joseph L. Dionne, four sections of papers are presented. In part I--Testing and the Basic Skills--P. Michael Timpane and Judith Shoemaker discuss federal activities, Edwin L. West, Jr. discusses the coordination of teaching and testing, and Amelia Ashe describes testing and teaching of writing competency. Part II--Testing and Information Services--includes Congressman Ted Weiss on national truth-in-testing legislation, William W. Turnbull on the use of information for particular purposes, and Michael H. Kean on why teachers dislike testing. Part III--Testing and the Teachers--included A Strategy for Uniting Testing and the Teacher, by Roy H. Forbes, Making Tests more Meaningful for the Teacher, by Wilson Riles, and Issues in Testing: The Perspective of Organized Teachers and Professors, by James G. Ward. Speakers for Part IV--Testing and the Publisher--are Oluf M. Davidsen, Lawrence W. Feinberg, and Ralph W. Tyler. Robert L. Ebel presents a summation, and proposes that testing for determing student achievement and testing to help pupils learn supplement each other. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

R. Bossone

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Proceedings

The Third National Conference on Testing: Uniting Testing and Teaching

March 13 & 14, 1980
Washington, D.C.

Edited by
Richard M. Bossone

Center for Advanced Study in Education
The Graduate School and University Center
of the City University of New York

Case 03-80
1980

MAY 2 7 1980

TABLE OF CONTENTS

		Page
OPENING REMARKS		
	Joseph L. Dionne	1
PART I	TESTING AND THE BASIC SKILLS: OPPORTUNITIES AND OPTIONS	
	Federal Activities in Testing: A View from the National Institute of Education P. Michael Timpane Judith Shoemaker	5
	Orchestrating Teaching and Testing Edwin L. West, Jr.	17
	Testing and Teaching for Competency in Writing Amelia Ashe	25
PART II	TESTING AND INFORMATION SERVICES: OPPORTUNITIES AND OPTIONS	
	National Truth-in-Testing Legislation: Reality and Myth Congressman Ted Weiss . . .	36
	Testing Information for Particular Uses William W. Turnbull	42
	Why Teachers Dislike Testing Michael H. Kean	54
PART III	TESTING AND THE TEACHER: OPPORTUNITIES AND OPTIONS	
	A Strategy for Uniting Testing and the Teacher Roy H. Forbes	63
	Making Tests More Meaningful for the Teacher Wilson Riles	70
	Issues in Testing: The Perspective of Organized Teachers and Professors James G. Ward	75

Page

PART IV TESTING AND THE PUBLISHER: OPPORTUNITIES AND OPTIONS

Testing: A Tool for Intervention
Oluf M. Davidsen 87

Testing and the Publisher: Can We Trust the Tests?
Lawrence W. Feinberg . . . 99

Achievement Testing and the Publisher: Opportunities and Options
Ralph W. Tyler 107

SUMMATION

Uniting Testing and Teaching: A Few Words in Summation
Robert L. Ebel 117

OPENING REMARKS

Joseph L. Dionne
Executive Vice President-Operations
McGraw-Hill

An outsider visiting our shores would report to his people that a great debate was taking place in America having to do with the role of testing in our society. Having noted the seeming acrimony of the participants, he would then admire the openness of a society which encouraged such deliberations. We, too, should celebrate our openness. There is, however, a time to rise above the tumult and shouting and to deliberate in a more tranquil setting. That is the purpose of CTB/McGraw-Hill and the Center for Advanced Study in Education of the City University of New York in bringing you, the nation's leading authorities in testing, together.

I would propose in your deliberations that you turn for direction to the core values of our society. Principally, there are two standards which I think might serve you well. The first is our continuing efforts to democratize society. We believe that, when possessing the needed information, our citizens will make informed choices. The democratization of information can be seen throughout our society. In the corporate world, we are seeing increased requirements for the disclosure of financial information, and in the political world we are seeing sunshine rules being passed to insure the openness of deliberations of our most significant institutions. So it is with the contents of testing.

Inexorably, all of us will be called upon to make public the contents of our testing program. It is unenlightened to try to blunt this movement. It would be far better through intensive research and development efforts to develop the technology by which we could rapidly change the items within tests and still have valid and reliable instruments. It is hoped that, through developments in latent trait theory or some other construct, a methodology for having a few items carry the requirements of credibility may be developed. Such credibility, married with recently developed database management technology, would provide us with the ability to create relatively inexpensive new forms of tests.

The second value to which you might want to turn in your deliberations is our belief, as a society, in progress through the individual. Schools will be judged by what they do for each child. Accordingly, we must individualize instruction. Diagnostic and prescriptive information is required. We need to develop information that leads to action. We need to understand which actions can be taken by the various participants in the educational process, including administrators, curriculum workers, counselors, teachers, students, and parents. For each piece of information to be gathered, we need to ask ourselves to what use it will be put. Will it facilitate learning? Will it lead to the reallocation of resources such as the amount of time spent on a subject, the methodology being used, the way in which students have been grouped, the structure of the subject matter, and the in-service education of teachers?

I have just two concerns as you seek solutions. The first is the potential liabilities of homemade tests. The movement toward democratization will inevitably be such that many districts will attempt to develop their own tests. Clearly, there are professionals present in each one of these communities who will want to see to the validity and reliability of the instruments to be developed. They should be heeded, for false-positive and false-negative information in the hands of teachers can be wasteful and, worse, harmful. Telling a teacher that a child has accomplished an objective when in fact he has not may rob the child of appropriate experiences. Telling a teacher that a child has not accomplished an objective when in fact he has mastered it can lead to repetitious boredom.

My second concern has to do with the explosion of the microprocessor technology now at hand. It can prove to be most helpful in developing management systems to the point where they can be a great aid to the teacher as well as to the student. But if they are employed in the way in which computer-assisted instruction was first utilized in the classroom--that is to say if they discourage interaction between teacher and student--they will prove to be ineffective and costly. If educators take the leadership in harnessing technology to meet their real needs, we can avoid the experience of having a technology in search of a nonexistent market.

Your deliberations are vital to the nation's interest. I know that they will be full and fruitful.

PART I - TESTING AND THE BASIC SKILLS: OPPORTUNITIES AND OPTIONS

FEDERAL ACTIVITIES IN TESTING: A VIEW FROM THE NATIONAL INSTITUTE OF EDUCATION

P. Michael Timpane
Acting Director
National Institute of Education

Judith Shoemaker
Senior Associate/Testing Team Leader
National Institute of Education

Federal policy toward testing has been emerging slowly over the last two years, largely in response to federal legislation (both proposed and enacted) as well as to concerns from our constituents. This emergence has been slow for two reasons. First, within the federal government, as elsewhere, there is a diversity of opinion over what role, if any, the federal government should play in relation to how tests are used. There are many individuals involved in setting policy in both the executive and legislative branches. Not all of these individuals share the same opinions regarding the federal government's role in testing. Secondly, the federal government is moving slowly in the area of testing to avoid any conflict over federal vs. local control of the schools. Federal regulation of testing or federally-sponsored development of tests could become backdoor approaches to federal control of curriculum. By regulating or specifying what is tested, the federal government could be specifying what is to be taught, in clear violation of our traditional position on local control of the schools. As a result, the federal government is moving slowly in the area of testing while trying to avoid any possible local-federal control problems.

Within the last two years there have been numerous events at the federal level regarding the federal government's role in testing. This paper will highlight some of the major legislative and programmatic events.

Legislative Events

Five legislative events characterize the interest of the Senate and House of Representatives in the area of educational testing. Although not all of the legislation described here was enacted, this list will give you an idea of the way the legislative branch has been thinking about testing and the federal government's role in its regulation.

1. U. S. Representative Ronald Mottl (Ohio) introduced a series of bills in 1978 to establish basic standards of educational proficiency. Initial versions of this bill called for a national commission to identify educational proficiency standards which would be used for determining high school graduation. After considerable negative reactions from the education community and the government, because of its threat of federal control, Representative Mottl's bill was amended considerably to provide grants to states waiting to identify educational proficiency standards and technical assistance in the area of achievement testing to states and local districts. This version was enacted as part of the Education Amendments Act of 1978 (P.L. 95-561).

2. The House Appropriations Committee, upon hearing that NIE intended to fund activities related to the development of tests with its 1979 budget request, immediately reduced NIE's appropriation

by $2 million to ensure that NIE would not be building a federal test. And Congress continues to show an interest in NIE's testing activities; for example, questions about testing arose again in both the 1980 and 1981 appropriations hearings.

3. Proposed truth-in-testing legislation, stressing the release of college admissions test scores, was introduced by U. S. Representatives George Miller (D-California), Shirley Chisholm (D-New York), and Ted Weiss (D-New York); and similar legislation was passed in the states of New York and California. To date, the federal legislation has not been enacted. Proponents of the bill stress its provisions for the rights of examinees of admissions tests to see their test questions and answers while opponents stress its costs and infringement. HEW was concerned about the proposed record keeping and clearinghouse functions called for in the bill.

4. The Department of Education Organization Act (P.L. 96-88), enacted in 1979, reasserted the policy of no federal control over local curriculum decisions.

5. Congress transferred the National Assessment of Educational Progress (NAEP) from the National Center for Education Statistics to NIE in a section of the Education Amendments of 1978. The language of that legislation made it clear that the design and development of a national assessment was entirely outside the hands of the federal government and placed squarely in the hands of an advisory policy committee. The legislation specifically

asked for a three-year report to Congress from NIE on the utility of NAEP data.

In addition to these recent legislative actions, there have been many federal, state, and local program developments in testing, highlighted below.

1. This nation has seen a tremendous growth in state and local minimum competency testing programs; of the 36 states that now have such programs, 16 tie test performance to high school graduation. This movement is unique in several ways: 1) it calls for more testing at a time when testing is increasingly controversial; 2) it is a movement begun by those outside the schools such as school board members and state legislators; and 3) it holds the danger that the student, rather than the school system, will be held solely accountable for learning failures, with consequences of failure assigned accordingly. In response to this rapid development in state policy, NIE initiated a descriptive study of these programs. Even at a descriptive level, we have had considerable difficulty in studying such a hot political topic. Our attempts at description were criticized for dignifying the movement by studying it at all, and for not developing impact information, which is a major concern of groups representing minorities. In addition, we have been accused of being pro-minimum competency testing for failing to describe alternate programs and we have also been accused of being anti-minimum competency testing by suggesting its possible adverse consequences. However, we still insist that such a

widespread movement needs study by NIE; it cannot be ignored.

2. In 1978, HEW sponsored the Conference on Achievement Testing and Basic Skills, largely in response to the prospect of federal legislative action in testing. The conference participants recommended that the federal role be one of collecting and distributing information on appropriate test use, conducting workshops and conferences on testing, and sponsoring research to improve testing. At the conference, Secretary Joseph Califano reaffirmed that the federal government should support, "rather than supplant," state and local testing efforts.

3. As a follow-up to the HEW conference, NIE held a Conference on Testing, Teaching and Learning to develop specific research plans in testing. The major recommendation from that conference was the need to improve the link between testing and learning. According to Ralph Tyler, the co-chairman of the conference, instruction is "the function least served by educational testing."

4. NIE also held eight regional conferences on testing, following the HEW conference, for teachers and administrators to discuss their concerns about testing. These two-day conferences included addresses on ways to link testing and teaching, small workshops on test use and development, and discussion sessions.

5. NIE established a research program on testing to provide direction to its expanding research, development and dissemination activities on testing.

As you can see by the preceding comments, NIE has vigorously entered the field of testing, even though we have avoided the dangers of federal control. Even so, you might wonder why did we enter this field at all, since it is so highly political and sensitive. NIE's mandate is to improve the practice of education equity. This charge is in our authorizing legislation, and it permeates all of our policy statements and program plans. All of our research activities are aimed at eliminating barriers for those who have not in the past had full opportunities to education; and all of our research should culminate in actions, programs, or dissemination that will improve practice.

Because testing is so much a part of the educational scene, NIE cannot ignore testing if it wants to increase equity or improve practice. We know that testing as it exists today will probably be around for a long time to come unless concerted efforts are made to improve it. We also know that tests have or can be used to bar educational opportunities for certain individuals and groups. Although we may get massacred in the fray between states, local districts, parents, advocates of affected students, and the test publishers, we are committed to improving the practice and the state of the art of testing.

In developing our research program on testing, we have had to make certain choices. We have chosen, for instance, to focus on the basic skills of reading, writing, and mathematics. The Teaching and Learning Program, of which the Program on Testing, Assessment and Evaluation is a part, has identified literacy as one of its major research areas. These skills areas also permeate the work of other areas within Teaching and Learning. We have

chosen research on testing as one way to improve the attainment of basic skills. Properly employed, testing can be used for early diagnosis of learning problems, for feedback and reinforcement in the learning process, and for the evaluation of learning outcomes. Within research on testing, we have also chosen to focus primarily on linking testing with instruction and on research to promote the development of alternative measures of performance and progress. We have chosen, finally, to look at the practices and policies of testing--its uses and abuses--as well as the methodology and psychometrics of testing.

The Testing Team has organized its current activities into two major areas: 1) current tests and testing practices, and 2) new methods of assessment and test design.

Current Tests and Testing Practices

As stated earlier, these activities recognize that current practices will probably exist for a long time; therefore, they focus primarily on reducing the misuse of tests through development and dissemination projects dealing with testing. Specifically, we sponsor two clearinghouses on testing which are designed to respond to the public's interest for information on testing: the ERIC Tests and Measurements Clearinghouse and the Clearinghouse on Applied Perforance Testing. These are both long-term projects supported by NIE. In addition to the eight regional workshops mentioned above, this year NIE will conduct five conferences for non-educators, including parents, lawyers, journalists, school board members, and legislative aides. The first conference will be held in Baltimore in

April for a regional meeting of Title I Parents. These conferences are being supplemented by a grant to the American Federation of Teachers which is conducting workshops on testing for its members.

The NIE Testing Team is also investigating the impact of testing on minorities, including studies of test bias, validity studies, and an investigation of the Law School Admissions Test. In our second study of minimum competency testing programs, we are focusing on their impact on minority students by conducting an educational forum which will present opposing sides on selected issues.

Finally, we are conducting in conjunction with the Office of Education a major study of how testing and evaluation information is used. This study examines 15 exemplary school districts and conducts case studies on test use.

New Methods of Assessment and Test Design

We also want to make a contribution to the improvement of testing and to the development of new methods of assessment. While we do not sponsor any test development work (lest it be construed as a National test), we do sponsor research on test design which we hope will lay the groundwork for others to develop new methods of assessing student performance.

At our conferences, participants have emphasized the need to improve testing as it is practiced today. They have claimed that many tests are unfair to minority test-takers and those with limited English-speaking ability because they result in decisions that are biased against these groups. This suggests that research is needed on factors influencing test

performance, factors associated with the test itself, and on the appropriate and inappropriate uses of tests for minorities. Participants have also claimed that many current tests are too narrow in content to assess the total range of desired educational outcomes. This suggests the need for exploring ways to assess the more difficult to measure areas, such as problem solving and decision making. And finally, recent reviews of testing literature indicate that despite changes in attitudes about the purposes of education, test development procedures have not changed much over the last 50 years. Reviewing accomplishments in standardized testing from 1927 to 1977, Oscar Buros stated that, "You might be surprised that so little progress has been made in the past fifty years . . . . Essentially, achievement tests are being constructed today in the same way they were constructed fifty years ago."

Most tests are still developed as norm-referenced tests, designed to rank test-takers along a single dimension. New advances in criterion-referenced and domain-referenced testing are just now beginning to influence test development. We have a long way to go to make tests appropriate for individualized instructional decisions. Our Conference on Testing, Teaching and Learning participants "envisioned, as a future ideal, testing merged into the teaching process, providing timely and rich feedback to the student, teacher, and others." We are not there yet.

During 1979 and 1980 our basic research grants program included a specific request for proposals dealing with testing and instruction. In 1979, we funded eight such proposals on the use of tests as teaching instruments. In 1980, we anticipate funding 3 to 5 proposals on how to

integrate test information with other assessment strategies to improve instruction. We are also sponsoring research on how test information is used in the classroom. Such studies have been undertaken for NIE by the Center for the Study of Evaluation at UCLA, Michigan State University, and Boston College. The later study explored the effects of introducing standardized testing in Ireland with special emphasis on how teachers used the results. Also under NIE's direction is a two-year contract with Michigan State University to develop materials for teachers on ways to interrelate test information with instructional decision-making. This study is co-sponsored by the Department of Defense Dependents Schools which will be used to evaluate the usefulness of these materials. NIE is also supporting the development of materials for teachers in the understanding and interpretation of standardized tests through a grant with the American Federation of Teachers. In the next two years, NIE will be engaged in the synthesis of these many projects which attempt to link testing, teaching and learning.

Future Activities

For the future, testing will remain an NIE priority, to be supported on a long-term basis primarily through the Teaching and Learning Grants Competition. This will ensure stable funding over the years for field-initiated studies in testing. These studies will also include small grants (less than $15,000) to encourage local districts and teachers to conduct their own research. These small awards should have a direct impact on practices at the local level.

We are also participating with the National Science Foundation on a joint, multi-year project to support research on computerized instructional components for mathematics, which should lead to the blending of testing, teaching and learning using new micro-processor capabilities.

Our future support for research on testing and instruction will focus on the use of tests as motivational and feedback mechanisms. We will also support the development of integrated assessment systems that use a variety of indicators to assess an individual's performance. Our emphasis will be on assessment systems for the classroom and on means for making decisions about admissions to higher education.

And we hope to study how tests are used in making decisions for selection and admissions of students into higher education. Although most of our previous work has focused on elementary and secondary testing, we realize that one of the major opportunities for creating barriers to educational opportunity occurs at the transition point between high school and college. We must learn more than we do about the extent to which tests are used to deny admissions for minorities; we must also identify exemplary admissions practices.

Educational testing as it is practiced today needs improving. We need better tests and better ways of interpreting test results. In most cases psychometric theory and statistical methods far surpasses practice and usage problems. In addition, we are beginning to understand the real politics and power of testing, and we are beginning to provide resources for the improvement of testing. In the short run, our

contribution will be the improved use of tests; but new methods of assessment are needed for the long run. We hope you will help us to reach both of these goals.

ORCHESTRATING TEACHING AND TESTING

Edwin L. West, Jr.
Superintendent of Schools
High Point, North Carolina

The conference theme, "Uniting Teaching and Testing," is particularly appropriate in light of the criticism many teachers and administrators have of programs which test the basic skills. In fact, a recent study on test utilization, conducted by the North Carolina Annual Testing Commission to provide information in the planning and improvement of testing programs, revealed a basic insight which, I believe, is common in many states. A random sample of 1,891 classroom teachers in North Carolina, with approximately equal percentages of first, second, third, sixth, and ninth grade teachers responding, revealed the following:

. Teachers at all grade levels saw testing reports and materials as helpful.

. Teachers with the highest levels of education and the most years teaching experience were most positive toward testing.

. As a group, however, teachers did not feel a need to receive training in test utilization and other important aspects of testing. Yet, almost half of the teachers sampled had never participated in any type of in-service training which linked teaching and testing.

In regard to this last finding, if we were able to completely ascertain the reasons educators do not feel the need for additional

training in test utilization, I believe high on the list would be that those of us in leadership roles within the testing arena have not, for various reasons, completely won the confidence and esteem of those teachers with whom we have come in contact. Among the variety of reasons are pressures of time, the rapid growth of the testing movement, and our own relentless tenacity to make our case and to "prove" that tests are the answer to all educational problems and can provide data which would transform group instruction into individualized instruction. Further, I suspect that in our zeal to support testing, we may, in fact, have exacerbated the problem by trying to dominate and overpower our opponents rather than striving to understand their concerns or problems.

This may be the essence of why there is criticism and a lack of desire on the part of many educators for additional training. And, until this problem is solved, we will never unite nor orchestrate teaching and testing. I propose that teaching and testing can be orchestrated. It requires, though, what I would call integrative unity, a concept which is not original with me. In fact, its origin can be attributed to the gifted observations and writings of Mary Parker Follett.

In essence, integrative unity is the bringing together of divergent strengths into harmonious direction. Basically, it is an approach which creates new solutions to accommodate, as fully as possible, the expectations of all individuals or groups. It requires a high order of inventiveness, perception and intelligence. It cannot be accomplished through executive order. Unfortunately, not many problems are approached this way nor can every problem be solved this way. Yet, I believe, we

could make a giant step toward orchestrating teaching and testing with such an approach. To make clear how this might be done, I would like to discuss a study that was recently conducted at the request of the North Carolina Annual Testing Commission on the variables related to pupil achievement in sixteen selected school systems in North Carolina--eight high achieving and eight low achieving districts.

The districts were matched in terms of selected demographic characteristics such as income and education levels of parents and of school district data such as per pupil expenditures. Research teams visited the school districts to verify data obtained earlier through responses to data-collecting instruments by personnel in the school districts.

The high achieving school districts had planned staff development programs which reflected the practice of integrative unity. The basics of integrative unity include the following: 1) relating the interests of the individual with the interests of the organization; 2) providing rewards which reinforce and encourage individual self-development within an organization; and 3) facilitating and fostering a dynamic and productive working climate. These basics may be applied in two ways to unify teaching and testing. One is the human dimension and the other, the organizational dimension. I shall speak first to the human side and then to the organizational dimension.

People grow from within. A test publisher, university professor, school principal or supervisor, cannot, in fact, develop anyone. However, they can challenge others to do their best. Herein lies, I believe, the

secret to the application of the three basics of integrative unity: gaining and maintaining the confidence of others. The high achieving school districts revealed that this can be accomplished through the application of four motivators. These motivators help to provide and maintain confidence at all levels needed to insure that teaching and testing are properly orchestrated:

1. Personal Gain. In any situation, many individuals will think or ask, "What do I get out of it?" It is the one question usually asked either consciously or subconsciously in most situations. If we are to orchestrate teaching and testing in the basic skills, we must demonstrate to teachers and administrators what they personally will gain from the testing process.

2. Prestige. Most people are motivated to participate in activities which help them receive recognition. Perhaps, we could better unify teaching and testing by recognizing, in some tangible way, teachers who have made significant accomplishments in using test data productively.

3. Imitation. Many persons have a need to follow the crowd and educators are no exception. If you visit many classrooms in the country, you will see imitation in action, for teachers use each other's workable ideas, learning activities, and other successful instructional techniques. I would suggest that if a teacher is employing a successful teaching technique related to test utilization, then we must use his or her success to stimulate others.

4. Convenience. Each of us in the education profession wants

things to be better and easier. If a new proposal appears easy, we will try it. Otherwise, we avoid it. If testing and teaching are to be orchestrated, then we must make tasks, procedures, and achievements easy rather than difficult. Sophisticated instructional management systems may not be the simplest or easiest to use.

You may be thinking what a simplistic approach to orchestrating teaching and testing. Or, those motivating concepts are too general and vague. I would argue that, on the contrary, to lead people to want to utilize test data in teaching in positive, productive ways, one must study their wants, needs and motivations on an individual basis, for individual thought governs group decision. Such action allows for the facilitation and attainment of educational goals while providing for satisfaction and growth for school system employees.

Thus far, I have shared with you only ideas or viewpoints related to the human dimension of orchestrating teaching and testing. I wish now to concentrate on the organizational dimension.

You will recall that one of the basics of integrative unity involved relating the interests of the individual with the interests of the organization. To accomplish this requires the presence of certain prerequisites on the part of the organization and, in particular, leadership within the organization. Research indicates that in high achieving school districts these organizational prerequisites existed: projection, patience, planning, prototyping, and politics.

1. Projection. School personnel must project or anticipate in advance the impact of change upon various constituencies and programs

within the system. By analyzing in advance those forces for and against change, problems can be projected in advance and measures taken to minimize conflicts which might accrue from implementation of the proposed change. This step assures that the impact or consequences of the change are examined from all aspects.

2. Patience. Inevitably, wherever people are engaged in working in a partnership on solving problems and practicing democratic processes, groups will go through the stages of dependency, conflict, cohesion, and interdependence. In addition, individuals within the group will undergo these stages at differing rates. It is, therefore, imperative that test coordinators, supervisors, consultants, university professors, and administrators understand this cycle and not overreact to individuals or groups trying to change their efforts.

3. Planning. Unless an organization knows where it is headed, what its constraints are, the resources it has available, and what evaluation data indicates, it will not be able to successfully orchestrate anything much less manage changes needed to effectuate orchestration. However, it cannot be assumed that individuals within an organization know how to plan. Therefore, it is incumbent upon leadership in the testing arena to be knowledgeable not only of testing but also of the planning process.

4. Prototyping. The space industry designs, builds, and tests prototypes of products and procedures prior to actual utilization of the model, product, or procedure on a wide-scale basis. Test coordinators and school administrators would do well to incorporate the same approach

prior to initiating and implementing wide-scale changes desired to orchestrate teaching and testing.

5. Politics. Any effort to unify teaching and testing must be examined within the context in which it is to take place. For this reason, testing personnel and administrators should examine the historical perspective related to the change and determine support or opposition involved in the attempt to unify teaching and testing in the basic skills. Doing this will enable the development of appropriate coalitions needed to effectuate change in a positive manner.

In summary, a review of the literature and research conducted in North Carolina reveals that those of us in leadership roles must be sure that the following human and organizational dimensions are in place if we wish to properly orchestrate teaching and testing in the basic skills:

- Motivating others to see personal gain, to feel prestige, to provide for imitation, and to assure convenience; and

- Being aware of organizational problems, having patience, planning, prototyping, and understanding the political consequences of actions.

The steps outlined are not panaceas but ingredients essential to the successful orchestration of teaching and testing. Through application of each step in a thorough and efficient manner, the prognosis for success is excellent.

REFERENCES

DeBruyn, Robert L. Causing Others to Want Your Leadership. Manhattan, Kansas: R. L. DeBruyn and Associates Publisher, 1976, pp. 45-48.

Smiddy, Harold. "Integrating and Motivating for Effective Performance." The Evolving Science of Management, Melvin Zimet and Ronald Greenwood, editors. New York: American Management Association, 1978, pp. 149-169.

Report on Test Utilization Questionnaire North Carolina Annual Testing Commission. Raleigh, North Carolina: North Carolina Department of Public Instruction, Division of Research, March 1980.

TESTING AND TEACHING FOR COMPETENCY IN WRITING

Amelia Ashe
Member, Board of Education
New York City

The decline in writing competency has been persistent over several decades. Precision, accuracy, and clarity are necessary for coherent communication. However, these qualities are in extremely short supply. Ask any professor of English, any business man, any editor, and even the deans of graduate schools where remedial English classes have to be imposed on would-be lawyers, journalists, and other professionals who have somehow avoided acquiring any acceptable degree of writing competency.

In the early 1970's, when public clamor and outrage about the decline in writing reached a crescendo, official attention focused, at long last, on the problem. Most of the attention came in the form of state mandates for the establishment of examinations to assure that "minimum competency" had been attained prior to the award of a high school diploma or certificate. In this climate of concern, a number of writing programs developed. Among these are the Bay Area Writing Project, in which teachers, working together, exchange information to develop expertise as writers and teachers of writing, and the Writing Proficiency Program by Dr. Richard M. Bossone, which grew out of a research experience that suggested instruction in writing is most likely to be effective when it is based on specific needs of students and when student progress is measured against specific objectives.

We are surely aware that teachers face a difficult task in teaching writing in today's schools. The conditions of teaching in urban schools

are such that teaching anything, much less writing, has become a dubious undertaking in recent years. The publication, American Educator, Spring 1980, contains an article titled, "The Conditions of Teaching" by Clifton Fadiman and James Howard. They state:

> Critics who heap blame on teachers for all the shortcomings of the schools are evidently insensible of the appalling conditions that too commonly constitute the disorder of the present day in American education. Public school classrooms are crowded with boys and girls of widely differing abilities so that reduced budgets may be met, and persistent notions of egalitarianism satisfied. The fractious children of a poorly disciplined society are not easily contained in classrooms, where their misbehavior is sometimes truly frightful. Teachers' hands are tied by the lengthening red tape of bureaucracy, while their tuition [teaching] is constantly interrupted for the sake of activities that entertain students and please the public, but educate no one.
>
> (Fadiman and Howard, 1980, p.6)

Such conditions are unpromising for the teacher of any subject, but particularly arduous for teachers of writing. Casual conversations with teachers of writing, as well as with the eighteen teachers in Dr. Bossone's in-service expository writing class, with whom I met, reveal the nature of the roadblocks involved.

We begin with an underprepared teacher. Even prestigious teacher-training institutions acknowledge that they have done little to prepare teachers to teach writing. Fadiman and Howard (1980) say: "So far as writing goes, teacher education has been out to lunch for years." We can assume that teachers trained in the 1950's, 1960's, and 1970's did little writing when they were students. A teacher who has had little

experience writing, either in his personal life or in his academic years, can hardly be counted on to impart to today's students the art of disciplined writing.

Of course, if poor writing instruction is as pervasive as we suspect, most students who appear in classes in any grade are likely to arrive underprepared for the writing task. Moreover, the distractions of mass-media communications: television, radio, telephone, the limping lyrics of rock music, and the staccato expressions of most magazines and newspapers often convince students that they are getting the message, despite the fact that they can neither write nor read with acceptable competency. Some authorities have gone so far as to argue that nobody needs to learn to read or write in today's world, giving students permission to avoid learning and the teacher permission to avoid teaching.

The combination of underprepared teacher and underprepared student reinforces the probability that the classic evasion of writing will occur. The teacher will tend to devote the preponderance of class time to the teaching of "literature." The teacher will first hope, then come to believe, that the elements of grammar taught in rote fashion will lead to mastery of elements of style as explicated in The Elements of Style by Strunk and White. When multiple-choice examinations take the place of essay writing, hardly anybody ever uncovers the fact that neither the student nor the teacher can be depended upon to write a coherent paragraph.

The most persistent and probably the most justifiable complaints about the conditions under which the teaching of writing must proceed concern

the nature of the teaching load and the grading of papers. In most urban schools, the teaching day consists of eight periods: five in which classes are met, one building assignment (i.e. study hall, patrol, or other administrative assignment), one preparation period (often used for "coverage" when a teacher is absent or for other school business), and a lunch period. In a few, usually private schools or suburban schools in wealthy school districts, teachers of English meet four classes a day with twenty to twenty five students in each--a sharp contrast to the teacher in New York City who faces five classes a day, with as many as thirty five or forty students in each class. In passing, I remind you that the work week for teachers has remained the same in the last hundred years, while the world has moved from a sixty hour week to a forty five hour week, to as few as thirty hours a week in the civil service and in some businesses and industries. As to grading papers, it is unrealistic to expect that a teacher can prepare writing lessons, teach writing skills, and grade papers for a teaching load of 175 to 200 pupils. How can the teacher give individualized instruction or personal attention to 35 or 40 students in each class? If care and concern about writing of an individual student is a consideration, then we are dealing in compromise and hypocrisy when we load teachers with the typical schedules, as described, and still expect constructive results.

Underprepared teachers, underprepared students, in combination with intolerable teaching loads, again reinforces the classic evasion of writing, and much worse, leads to hopelessness. When teachers sense that

they have been programmed for failure, many tend to give up. Assuming a modest two writing assignments a week, how can a teacher correct, i.e. read carefully, comment critically, and evaluate 350 to 400 papers a week? At only five minutes per paper, the requisite time would be about fifteen hours a week, or two long work days in addition to the five in school.

Another justifiable complaint has to do with the imposition of non-teaching chores, endless class "equalizations" that may take as much as a whole of the first month of a semester, and a dearth of supplies as basic as a scarcity of paper. If we add these to the "appalling conditions" outlined by Fadiman and Howard (1980), as quoted above, which cited heterogeneous grouping and interruption of extraneous activities into the teaching day, then we begin to understand what brave hopes lay behind Dr. Bossone's attempt to teach some New York City teachers of English to become effective teachers of writing.

Even before 1967, when I first met him, Dr. Bossone, now University Dean for Instructional Research and Professor of English at the Graduate Center of the City University of New York, was an advocate of training English teachers to teach expository writing. As a member of the New York City Board of Education, to which I was appointed in 1974, I was sympathetic and encouraging when Dr. Bossone came forward with a proposal to develop A Strategy for Coping with High School and College Remedial English Problems. As a result of my support and his investigation,

Dr. Bossone developed The Writing Proficiency Program which includes tests and instructional materials in a diagnostic-prescriptive mode for the teaching of expository writing, grades 9-13.

In the fall of 1979, Dr. Bossone conducted a special in-service program for eighteen teachers from twelve different New York City high schools. These teachers were enrolled in a graduate level class at CUNY designed as a pilot project to test out his writing program. The eighteen teachers in the graduate course, in turn, taught approximately 600 students enrolled in English classes in our high schools. About seventy percent of the students were Black, Hispanic, or Oriental. Most came from low socio-economic families. About sixty-five percent of the students enrolled worked at outside jobs, typically from fifteen to twenty hours a week. A large number of the students reported that they did not like English at all. Some said they liked English "somewhat." Approximately seventy percent of these students were classified as "below average" by their teachers; the remaining thirty percent of the students were classified as "average" in academic ability.

Fifteen of the eighteen classroom teachers of English taught eleventh grade classes; two taught cross-graded eleventh and twelfth grade classes, and one taught twelfth grade. Fourteen held master's degrees. The others held baccalaureate degrees and offered a mixed-bag of courses "equivalent" to a master's level. The majority of the teachers believed they did not have the proper training as undergraduates or graduate students to teach writing. Their teaching

experience ranged from a low of three years to a high of twenty-two years. Typically, they had eight years of teaching experience.

Originally, almost 600 students were enrolled in these classes, but only 287 students took both pre- and post-objective exams and wrote pre- and post-paragraphs and essays. Contributing to this marked attrition were normal problems of attendance, class changes in "equalizations," and the serious drop-out rate in New York City high schools. Added to all this, the eleventh grade is particularly vulnerable, by age of the student and the student's inclination to give up if ultimate prospects of graduation begin to appear remote. Given the characteristics of the student population in this pilot project, the previous lack of specific training of teachers, and the contraints inherent in the conditions of teaching, the major findings appear to be truly remarkable.

I shall spare you details of raw scores, pre- and post-test. Let me summarize: seventy percent of the students who took both pre- and post-objective tests showed a net gain in the number of objectives mastered from the pre- to the post-test. Eighty percent of the students showed significant improvement in their ability to write paragraphs and essays. In general, the greatest gains were shown in the section on sentence structure and paragraph organization. The greatest improvements reflected in the writing samples were in paragraph unity and coherence and in essay organization. Improvement was also reflected in sentence structure. These gains were shown after the

students had received only ten weeks of instruction, or one hour and forty minutes of instruction a week, or a total of approximately seventeen hours of instruction in the fall semester, 1980.*

Fourteen of the eighteen teachers in the pilot project reported that the course they took with Dr. Bossone definitely helped them to become better teachers of writing. The other four believed the course helped them "somewhat." In critiques of their experience in this program, one teacher wrote:

> The Writing Proficiency Program is a step in the right direction. Classroom teachers are trained in the techniques of teaching writing in a very systematic, logical and comprehensive program. The program enables the teacher to diagnose student writing, prescribe a plan of remediation and evaluate progress. In requiring that the teacher write too, the program dramatizes for the teacher the difficulties inherent in a writing task and the rewards.

Another wrote:

> The class with Dr. Bossone has been a great help to me, professionally and personally. The students in my classes enjoyed using the materials, and derived much profit. The breakdown of individual scores, strengths and weaknesses spurred them to improve and achieve results:
>
> 1) Dramatic improvement in written expression, sentence structure and mechanics.

---

* Supervision in the course of the previous research study--A Strategy for Coping with High School and College Remedial English Problems determined, by actual observation and careful count, that only eighteen to twenty two minutes of any given English period were devoted to activities that might even remotely be considered instruction. This finding was also confirmed by teachers in this pilot study.

2) Improvement in student understanding of the principles of organization.

3) Improvement in student understanding of topic sentence, other concepts, specifically, main idea and supporting details, going from the general to the specific.

4) I have noted improvement in my own writing.

Yet another:

> I was very happy at the end of the term when I saw the results my students produced on the post essay test. All of my students improved this term, as judged by objective graders. Their improvement is the thing which in the last analysis made this course so worthwhile.

And still another:

> The major problem in teaching English in a New York City high school today is the total unreality that possesses our Central Board, the bureaucracy and the politicians.
>
> The surrealism that we labor under are classes that are in constant flux; numbers that can reach forty; no supplies (paper, pencils, and limited books); a huge range of talents and attendance on the part of the students. Moreover, reading and listening to the media as they report the politicians' criticisms and suggestions is demoralizing. Worse yet, it is hypocritical.
>
> Finally, to the credit of the Board, Dean Bossone's course at least gave some recognition to giving a program leading toward a path of solution for the teaching of writing in the high school.

I cannot speak for the politicians, nor for the "bureaucracy," but I can for myself as a member of the Central Board. Happily, the teacher-critic quoted above has at least a minimal perception that the "Board" did cooperate, or give some recognition to the problem, in encouraging

Dr. Bossone to test his writing program by training New York City teachers in its use. "The path of solution for the teaching of writing in the high schools" is obviously the signal success of the Writing Proficiency Program in its impact on both teachers and students.

Too little, too late--perhaps. But from feeble beginnings we inch toward progress. Browning's dicta apply: "A man's reach should exceed his grasp . . . . ," and "The best is yet to be."

PART II - TESTING AND INFORMATION SERVICES: OPPORTUNITIES AND OPTIONS

NATIONAL TRUTH-IN-TESTING LEGISLATION: REALITY AND MYTH

Congressman Ted Weiss
20th Congressional District
New York City

Today, a few testing agencies exercise enormous power over the educational and occupational future of millions of Americans. The testing industry, despite its influence on vital aspects of individuals' lives, operates in an unaccountable fashion. Little or no input in the testing process is received from interested scholars, test-takers, large segments of the education community or the general public. The President's Commission on Privacy in 1977 observed about testing and data assembly organizations: " . . . although such organizations deal directly with individual applicants, and collect and process mountains of information about them, they are less accountable to the individuals on whom they keep records than any other type of record-keeping institution in higher education."

Perhaps it is because of the vast power of the testing organizations that standardized tests have acquired an aura of omniscience. They are generally regarded by test-takers and by some academic evaluators as the single most important indication of a person's ability to perform acceptably in a particular academic setting. Yet, test results can literally change the course of a person's life. All of us have heard about "cut-off scores" below which an individual is not admitted to a school, and we are likewise familiar with the status accorded someone who

has scored 700 or higher on a standardized test. A poor score on one of these examinations often inhibits or stifles completely an individual's academic ambitions.

In addition, tests may reflect biases that result in lower scores for certain geographic, racial, and income groups. In testimony before the House Subcommittee on Elementary, Secondary, and Vocational Education, sociologist James Loewen, from the Center for National Policy Review at Catholic University, demonstrated that there was a high correlation between PSAT test scores and income as well as between test scores and geographic location. According to Loewen, a person from an economically disadvantaged state, such as Mississippi, is more likely to do poorly on the PSAT than a person from a prosperous state such as Connecticut.

To help avoid such injustices I, along with Representatives Shirley Chisholm of New York and George Miller of California, sponsored the Educational Testing Act (HR 4949), commonly referred to as national truth-in-testing legislation. The purpose of this Educational Testing Act is to encourage greater accountability to the public and facilitate openness and innovation in the testing process. It will enhance the opportunity for scholars outside of the ETS-ACT circle as well as the general public to have input into the testing process. As the noted scholar in testing, Oscar Buros, stated: "If detailed information were made available simultaneously with the publication of tests, not only would educators and testing specialists be in a better position to evaluate the tests and to interpret the results more intelligently, but publishers would be impelled to construct better tests, pointing out both their strenths and weaknesses." Buros' observation suggests the

legislative solution embodied in HR 4949. The bill is essentially a disclosure measure, compelling testing organizations to provide more information on tests and testing procedures.

The accuracy of these tests and the way in which they are prepared are not publicly known. The testing companies treat these exams like highly classified documents, withholding background information from both test-takers and reputable researchers. Testing organizations conduct their own studies on the validity and accuracy of their exams on a regular basis, but the findings of their studies are not available to "outsiders." Buros termed these efforts of the testing organizations "quite unsatisfactory." During the course of the numerous hearings held on HR 4949 last year, many researchers in testing confirmed this observation of Buros and endorsed the bill as a way to facilitate greater access to research materials.

The Educational Testing Act would accomplish its purpose in four basic ways. First, it would require that testing agencies include a minimum amount of information about the test's purpose and methodology with the registration form for the exams. For example, a testing agency must state in clear language the aim of the test, the skills being assessed, the exam's margin of error, and other related information. While most testing agencies currently provide an explanatory manual along with the registration form, the extent and quality of information provided in each manual varies from test to test and agency to agency. This provision would simply ensure the distribution of uniform information.

Second, HR 4949 would also require that testing agencies share with the Secretary of Education the results of any studies or statistical analyses they prepare on their tests. These findings would become part of the public record and be accessible to independent researchers in the field of testing.

Third, the legislation would require the testing agencies to provide the Secretary, within 30 days after the results of a test are released, the questions used in calculating the raw score for an exam, the corresponding answers, and all criteria for translating the raw score into the score that is reported to an institution. This information should eliminate some of the mystery surrounding the standardized testing procedures. It would also enable independent researchers to review the questions and answers used in specific exams and suggest possible improvements in test methodology.

In addition, this section of HR 4949 would make the tests more relevant to a person's own educational objectives by permitting test-takers to obtain copies of their answer sheets and the correct answers from the testing agency. The test-taker now has only the most general knowledge of individual strengths and weaknesses because scores are broken down into general categories such as verbal or quantitative. However, a detailed accounting of results in each area will assist the test-taker in understanding where to concentrate efforts on improvement.

I must emphasize that the legislation requires only the release of questions and answers used in preparing the raw score. The bill also exempts from this provision those tests administered to fewer than 5000

subjects during a testing year. This particular requirement should, therefore, not cause any significant increase in expenses to the testing agency or, by extension, to the test-taker.

Finally, the legislation stipulates that testing agencies must file information with the Secretary of Education about their fee structures and expenditures. This provision would further ensure the public accountability of the testing industry, large segments of which enjoy tax-exempt status as well as federal contracts.

Many critics of our bill have greatly overstated the scope and impact of the legislation. In no way does the bill demand alterations in the examinations or in admission standards for academic institutions. The legislation, in fact, strengthens the reliability of standardized testing by helping to demystify the entire process and to establish a method of cross-checking scores. Further, our legislation aims to eliminate biases in the selection of questions, not by government decree, but by enabling outside researchers to review test preparation criteria and to conduct an informed dialogue on test methodology.

The testing companies claim that the requirement of public disclosure of test contents will necessitate preparation of many more questions, thereby weakening test reliability. This claim is invalid because a large portion of the questions on many of the major tests are brand new. An official from the American College Testing Service, for example, testified at congressional hearings that currently on four out of five ACT test administrations, one hundred percent of the questions are new. In addition, HR 4949 protects procedures for

maintaining test quality by its exemption of test items that do not affect the raw score.

Some critics also assert that revealing questions and answers will be a boon to coaching schools for the well-to-do. But these schools already exist, and their clientele is almost exclusively from affluent families. HR 4949 is likely to make effective coaching readily available to students from all economic backgrounds since practice tests will be widely distributed.

One of the most ironic criticisms of HR 4949 is that it would lead to a "nationalized" admission test. In a sense, testing is already nationalized--there is only one LSAT, one MCAT, one GRE for the entire country.

The Educational Testing Act, I believe, is a vital and overdue reform essential to improving the testing process. It addresses the need for improvement in tests which in 1977 Mr. Buros marked as showing "relatively little improvement" over the 50 years he had observed them. It will, at least, bring us further toward fulfilling what the NIE-sponsored Conference on Research on Testing (August, 1978) termed a "fundamental human right and a necessity"--the disclosure of test items and the students' answers.

TESTING INFORMATION FOR PARTICULAR USES

William W. Turnbull
President
Educational Testing Service

Every year or so, in my family, we take solemn vows to do something about the way in which printed matter seems to be getting the upper hand in our lives. We review and even prune back our subscriptions to magazines--only to find that our family reading list responds to pruning just the way our shrubs do: by developing two new sprouts for every old one lopped off. You may have had the same problem. The guardian spirit of this age of the information explosion must be Hydra.

Inevitably, our times are characterized as the era of "information overload." We seem suddenly inundated with information--so much of it that we can no longer cope with it all. People have begun to suggest "coping strategies," such as "selective apathy" to particular aspects of the world's affairs, so that one conserves valuable time for one's own concerns.

Education has its own information overload. The research reports, the news accounts, and the statistics about education add up to much more than anyone can keep up with. And yet, I am not at all sure we have the information we need to answer the critical questions about either the individual student or the functioning of the educational system. Moreover, I am convinced that we are not as well prepared as we should be to

interpret that information if and when it becomes available to us.

My thesis, then, is that we need to think of making information available for very particular purposes. This would consist as much of selectively reducing information as of adding to it even though we may well need more information of some kinds than we have now. And then we need to know its meaning and its limits in order to convert raw information into a legitimate basis for decision.

I believe that testing can and does provide some of the most basic intelligence needed to improve education if the information is sound and we know what it means. I believe also that gathering this information can and should be streamlined to effect better investment of school time.

Since the theme of this conference is "Uniting Testing and Teaching," let us start with the use of tests to provide feedback to student and teacher at the level of the specific learning task. Surely, this is a basic element in the teaching-learning process. A style of study requiring an active response, reinforcing right answers and keeping the student's attention concentrated on Step A before moving on to Steps B and C, fits what we know about learning. It properly shifts the action--the responsibility--as far as possible to the learner, who is in charge of the learning, while giving the teacher the best possible basis for letting students move at their own pace without losing track of their progress or the

ability to help each individual student when he or she needs encouragement and help. The intelligent use of new technology may greatly enhance the spread of this approach to education, so long heralded and, so far, so little realized in practice.

A prime requirement for such an approach is the development of the course and test questions as integral and inseparable parts of a single whole. One of the best opportunities for improving education lies in the creation of more and more effective teaching-learning materials in the form of curriculum-embedded test development through a close collaboration of specialists in curriculum and testing.

In many situations, the author of the course, or a particular segment of it, is the teacher. Here the same rule applies: the test questions should be developed as integral parts of the teaching material. This means that the teacher has to be as good at the skills of writing test questions as at those involved in outlining the course of instruction. Therein lies a problem because few teachers learn test-writing skills as part of either their pre-service or their in-service training. Much of the value of uniting testing and teaching will be lost if the two elements in the union are not matched in quality.

Test questions that are embedded in the course, that carry part of the instructional message, and that allow teacher and student to keep track of individual progress are at the heart of enhancing learning through what is usually called "formative" evaluation; at this point the aim is to promote and advance learning as work in progress rather

than to reach a "summative" conclusion about it. Ideally, such questions may also serve to direct the student forward or backward in the course material, depending on whether the answers show complete understanding or a particular misconception on the part of the student.

The information provided by a student's answers to curriculum-embedded questions can in theory be used to produce evidence of the student's mastery of any course unit and indeed of an entire course without further testing, except for one human problem. A student may be able to answer a question correctly upon first having learned it but not retain the knowledge or skill until the end of the course. It is quite possible, however, to produce a separate course exam, for later administration, that tests knowledge of the key concepts in the instructional sequence, reinforcing and consolidating earlier learning, as well as helping to sum up the student's cumulative progress.

The information resulting from the embedded tests is of obvious importance to both teacher and student but, for the most part, to nobody else. The information provided by the course exam is also very much their concern. Accordingly, it is essential that the teacher play the key role in selecting or constructing the course exam. Lack of such involvement is the surest way to insure that the results will be ignored.

Up to this point, I have been talking only about information describing the individual student. Inevitably, however, the teachers, the principal, and others need to analyze the results for the whole class and for the school as a whole in order to see that the system is working well.

It is at this stage that the school may want to use an external test, in addition to the course exam, as part of the summative evaluation process to compare student's achievements with others in comparable settings. If this step is taken, it is obviously essential that the external test be chosen carefully to fit the objectives of the course and that the teacher play a key role in its selection.

From here on, the information needed is for wider and wider audiences: the district superintendent, the school board, and the community. The point here is that with cooperative planning throughout a district and, indeed, across a state, elements of the same data should be designed to meet classroom instructional needs, end-of-course grades, and most other administrative needs to reduce redundant testing. I believe that a study of alternative models for accomplishing this purpose, including an integrated, statewide information-handling system, could pay large dividends in time and information.

Where do comprehensive nationally available survey tests selected by the school district fit into this picture? I believe they do have a place because they can sum up the cumulative knowledge and skill that the student has both acquired and retained over a span broader than the course. They also permit comparisons of students and classes from year to year and from school to school if equated editions of the test are available. This is especially important for people outside the school system, such as college admissions officers who need comparative information that

crosses district boundaries and state boundaries; and by the same token, it is important for students who need to know where they stand and for parents and counselors who want to be of help.

One possible result of such a carefully designed blend of assessment techniques to create a comprehensive school-based program in most schools may well be a reduction of emphasis on exams wholly external to the school, such as exams for college entrance. Let me quote one section of an article I wrote for the College Board Commission on Tests and which appeared in Science a dozen years ago:

> To my mind, the ideal system is one in which the information needed not only for instruction and guidance but also for college admission and placement is gathered, organized, and used in the school. By the time a person reaches college-going age, he has obviously built up a very substantial record of accomplishment, differentiated according to the abilities and interests he has already displayed in school. The information gathered in the relatively brief time devoted to the external examination should, under the circumstances, be redundant.
>
> It seems to me that we should work toward a system for collecting, on a regular basis, information about each child's performance as he passes through the school system--a record of his accomplishments in terms of grades and standardized tests, his interests, his extracurricular activities and so on. We need to find a way of organizing this information efficiently, of expressing it in an unambiguous language that can be communicated to other people and manipulated rationally, of storing this information and of retrieving it rapidly for use in student guidance within and between educational levels, including grades 12 and 13.

If the several elements in a school's testing program and in a statewide program are carefully articulated, it is a fair guess that the quantity of the information gathered will decline and the quality will rise.

There is an important caveat to the assertion that quality will rise. Ultimately, information has quality only if it is interpreted with skill and understanding. My observations lead me to worry about the lack of specific training of the people who use test information: lack of training in preparing tests or in interpreting the results. This comment about lack of training applies whether one is talking about classroom exercises, nationally standardized objective tests, or essay exams.

Tests and their interpretation are obviously the subject of heated debate across the country these days. Some of the assertions made in the public press are inevitably expressions of popular misconceptions rather than statements of fact or expert opinion. The burden falls on teachers and administrators not only to sort out the elements of these controversies for themselves but also to clarify the issues and facts for concerned, uninformed people. The unhappy truth is that relatively few teachers or administrators today receive so much as one course in testing as part of their preparation. As a result, a majority of teachers and administrators are less equipped to deal with this subject than its present prominence demands.

The subject of testing is not particularly difficult or abstruse, particularly if the emphasis is on measurement concepts and on making good tests rather than on statistics which seem to turn so many teachers off. The quality of the use of test information would be enhanced greatly if such a course were required as part of teacher education programs and made attractive enough to teachers so they would want to take it. In today's climate, is is difficult for me to understand why it is not

accorded this priority in teacher education.

Equally urgent is the need for school administrators to receive instruction in the basics of test interpretation. While the teacher needs to concentrate on gathering and interpreting information about the individual student, the school administrator needs to understand fully what meaning to attach to data for a whole class, school, or community. Think of how much time and confusion over media interpretation of test data could be saved if each school had at least one person, preferably the principal, with a modicum of expertise in test interpretation. Courses tailored to that purpose should be required and made widely available on a pre-service or an in-service basis. Let me add a similar plea for courses for people at all levels who use test information, as well as grades and other data, in admissions. It could perhaps be made standard practice to insist that at least one member of any admissions committee have demonstrable expertise in the strengths and limitations of the kinds of data considered by the committee.

I have been speaking about the need for teachers and administrators to properly interpret test scores. Earlier, I mentioned the equally important need for students to understand information, including test scores, that could help them and others make decisions about their futures. The recent debates about the use of tests have suggested to me, and I am sure to others, that we have not done as good a job as we could in giving students full understanding of their test performance and the meaning of their scores. This is an area in which test developers, publishers, and program sponsors can and should take the lead. The written communications

to students and the nature of the reports they receive are important and can carry basic information. It is important that knowledgeable people, such as well-trained counselors, teachers, and administrators be available in the school to help in this communication with students.

The interest of many people in extending and deepening students' information about the test scores and their ability to interpret scores appropriately has found expression over the past year or so in proposed legislation. Some of the suggested legislation, such as that proposed recently in New Jersey, would apply to any and all standardized tests at any level and for any purpose. Legislation actually passed in New York State and a bill introduced last year at the federal level by Congressman Weiss would apply, however, only to programs of testing for admissions at the undergraduate or graduate level.

The aim of increasing student awareness and understanding is laudable, and I applaud it. For reasons that have been well covered in the press, however, many testing people believe that most of the specific means to that end embodied in the proposed legislation are superfluous and would do a great deal more harm than good. In particular, the requirement that each test be published soon after it is given and, therefore not available later for re-use, threatens to lead to substantial increases in the cost of tests and to a great decrease in the number of times per year they are offered. Enactment of a law in New York with such a provision has already led to the disappearance of 20 out of the 26 admissions test programs formerly offered in that state and, for the other 6 programs, both a curtailment of service and an increase in fees. New York is now holding

hearings on possible amendments to its law, which, I believe, was enacted too hastily on the basis of an incomplete understanding of the facts.

If the aim of expanded test information is desirable but the proposed legislation is undesirable, what can we do? In the realm of testing for admissions, where much of the controversy has been concentrated, this question is currently under very active discussion with the aim of finding and instituting voluntary solutions that avoid regulatory legislation in general and the unfortunate side-effects of higher fees for less service in particular.

Several educational groups that sponsor admissions testing programs have indicated their intention of providing more complete information about the tests for which they are responsible. This intention was expressed clearly in a document entitled Public Interest Principles for the Design and Use of Admissions Testing Programs issued late last year by leading people in program sponsor groups, such as ETS, which works with all four. They have proposed several propositions for discussion, elaboration, and early implementation, including such statements as the following:

1. that each prospective examinee should be able to receive a full-length sample of each test, similar to the one he or she will take, with the intended answers and with instructions for self-administration and self-scoring.

2. that for tests given to a sufficient number of students annually to support the cost, at least one operational form of the test should be published periodically, in addition to the regular sample. A specific schedule of publication should be designated for each program.

3. that non-technical information about the testing program should be furnished routinely to test takers, users, and the general public. It should include a description of what each test measures, the error of measurement, how the scores are intended to be used, and a summary of the validity of the scores for the intended uses.

4. that a technical publication should provide information on the same topics in sufficient depth to permit professionals in the field to assess the evidence and the accuracy of the non-technical summary.

5. that tests, together with the procedures for scoring them and reporting the results, should be designed to provide test takers with as much useful information as may be feasible about the specifics of their performance on the tests.

6. that the program sponsors should formulate, maintain, and publish widely principles of appropriate use of scores and other test information derived from testing programs and should be alert to and actively discourage misuse.

These proposed actions are not all new for any single program represented by the signatories. What is new is their formulation for discussion as operating principles or guidelines for this important set of external admission programs. They will, I believe, enhance student understanding without crippling the operation of the programs and thereby reducing the valid information available to students and institutions alike. It would be useful to consider formulating comparable principles and implementing guidelines for the instruments used in a school's own testing program.

In sum, I think there are opportunities for better information services about tests and for better use of information provided by tests. The two are inseparable. The information provided by tests internal to the school and used for teaching purposes should be aggregated with external test data to serve administrative purposes while cutting

redundancy. But the value of testing for all purposes can be no greater than the skill that goes into the development of the tests and the use of the scores. That skill is now far below where it could be and should be. Extending the information about tests, scores, and increasing the expertise in the interpretation of test information by all people involved deserves a high priority in the 1980's.

WHY TEACHERS DISLIKE TESTING

Michael H. Kean
Executive Director, Office of
Research and Evaluation
School District of Philadelphia

Few, if any, of us have ever stopped to think about why teachers dislike testing. We, in testing, have established an adversarial relationship with teachers and attempt to deal with testing on a "thou shalt" basis. Let me suggest to you that in order to appropriately deal with the problem, it must first be carefully and honestly examined. Such an examination might yield the following basic reasons as to why teachers dislike testing: tests are too difficult to understand, tests offer negative perceptions, tests do not measure what they purport to measure, tests are not related to classroom instruction, and tests are too costly and time consuming.

1. Tests Are Too Difficult To Understand

Many teachers have little or no prior exposure to tests and measurements. Most undergraduate courses in teacher preparation do not include a course in testing. Even at the graduate level, if such a requirement exists, it is likely a single statistics course, often taught by someone more interested in theory than in practice. To many teachers, this exposure to statistics is a negative experience and one which they associate with testing in general, particularly when test results are presented in a technical fashion.

2. Tests Offer Negative Perceptions

Many teachers believe tests are harmful to a student's self image. This belief is further reinforced by the fact that a great many teachers had negative experiences associated with testing when they were students; and it is further reinforced by the media which tends to capitalize on antitesting or negative information about students' performances.

Tests are also perceived as being a threat to teachers' images. Many teachers fear that tests will be used to rate or evaluate them. Also, many teachers view testing as an intrusion into the sanctity of the classroom. This need not be so. A resourceful teacher can make use of test information to maintain the autonomy of the classroom by designing specific instructional strategies based upon test results. The fear of intrusion, I suspect, is related to the fear of public disclosure.

3. Tests Do Not Measure What They Purport To Measure

Many teachers believe testing represents an unrealistic measure of student effectiveness and/or competency. Nationally normed standardized testing programs represent but a single measure; hence, it is unrealistic to assume that their results can be directly related to student effectiveness. In addition, such tests are administered at varying points of time in the school year and represent varying degrees of pupils' exposure to a particular teacher. Most test scores represent but a single point in time, and unless trend data are presented, it is impossible to compare growth.

In addition, the question of test norms is typically raised in the context that national norms do not reflect, for comparative purposes, certain types of school districts. Questions are most typically raised in urban areas and in rural poverty areas, where scores tend to be systematically lower than the national norms. As a result, there has been a call from some quarters to develop separate urban norms. Though such norms may well have advantages, it has been virtually impossible to provide comparative information among the largest 20 or 25 school districts in the country. This is due to a number of factors, including the population tested, the time of year testing takes place, the instruments utilized, the metric used to report results, and the availability of those results. When in-level and out-of-level testing are also factored in, the problem becomes even more complicated.

4. Tests Are Not Related To Classroom Instruction

Teachers often complain that test information is simply not used or not used properly. This stems from the fact that teachers seldom have an opportunity to view what, if any, use is made of test results. If the school principal, who is still viewed by many as the instructional leader of the school, fails to make use of test results, can we legitimately expect the teacher to do so? The same holds true for the curriculum director, the director of reading, or the director of mathematics. And as pointed out earlier, teachers are rarely taught how to make proper use of test results, and should they express a desire to do so, assistance is

rarely available.

Also, teachers view many tests as being too difficult for their students. This is particularly true in urban areas where scores are typically below the national norms and substantially below the results reported in the nearby suburban districts. As a result, debates have been spawned as to the wisdom of the use of in-level vs. out-of-level testing. Unfortunately, the rationale for testing out-of-level is typically tied only to a desire to raise test scores, not to a desire to improve instruction.

In addition, teachers view these tests as being too difficult to integrate with instruction. The testing program, particularly, the administration of tests, is often not integrated within the instructional program. Results are typically not provided in a way to enable teachers to use them to improve instruction. Even when item analyses are provided, they are rarely accompanied by a training program to help teachers interpret and make use of such data in tailoring their teaching to the needs of the individual student.

Then, too, teachers believe many tests do not correlate with the curriculum. Very frequently, teacher input is ignored in the selection of tests. As a result teachers often question the curriculum validity of the instruments they are required to utilize. This becomes an even greater problem given increasing levels of curricular autonomy and flexibility.

Finally, many teachers think tests are not sensitive to a single year's instruction. This is true because teachers are often viewed as

being solely responsible for a student's score. Yet we all know nationally normed standardized tests are, of course, cumulative measures and relate not only to that which a student has learned in school in previous years, but also to a variety of other extremely powerful, non-school influences.

5. Tests Are Too Costly And Time Consuming

Lastly, many teachers believe tests cost too much. This is simply not true, though common folklore has enabled the myth to survive. The typical cost of mounting a nationally normed standardized testing program represents a tiny percentage of the typical instructional budget. And this cost can be further reduced by careful planning, in-house scoring, and other economies.

Further, many teachers believe testing takes too much time. Although a mismanaged testing program can, in fact, represent an unconscionable waste of time, this need not be the case. The time wasted is not the actual test-taking time, however, but the time taken to implement the program, and collect and process the instruments. This can be easily remedied, however, by a carefully developed system which provides for the pre-slugging of answer sheets, differentiated distribution, centralized editing, the provision of class lists, etc.

Suggestions

Though many of the above reasons for teachers disliking testing may not be, upon detailed examination, valid, each one is a legitimate concern in that each affects the way teachers feel about testing.

How can we best deal with these problems? We could take each one and respond to it. Such responses, however, are often viewed as more defensive than constructive. Instead, I would propose a series of general suggestions to deal with these problems. Consider, for example, the following:

. A thorough explanation needs to be developed concerning what tests and testing can do. Similarly, an explanation must be provided concerning the limitations of testing (what tests cannot do).

. Technical language must be avoided at all costs.

. The results of testing programs should be communicated in a timely, useful, and understandable fashion.

. Alternate test score reporting forms should be developed and tailored to the needs of specific clients and client groups (e.g., parents, the media, etc.).

. The teacher should be provided with a series of different reporting forms which might be tailored to different uses (e.g., parent conferences, instructional planning, program development, etc.),

. Teachers should be provided with extra supportive services related to the testing program. The school principal, assistant principal, testing program coordinator, or the district testing assistant should all be available in a supportive/consultive/ instructive role before, during and after the testing program.

. The appointment of an individual to function as a school

district testing ombudsman (on an ongoing basis) might help to identify specific areas of weakness in the district's testing program and delivery system.

. The testing delivery system must be made unobtrusive. Teachers should not be required to perform clerical tasks.

. All questions should be answered, even those of a technical nature, though special care should be taken to answer technical questions in non-technical language.

. Resistance to the testing program should be expected, and those responsible for planning the program should be prepared to deal with such resistance when it occurs.

. Teachers should be given instruction as to various kinds of tests and exactly what the various tests measure.

. The testing program, particularly its output, must be directly linked with the instructional process. The relationship between testing and teaching, including its cyclical, decision-oriented relationship, must be stressed.

. Teachers should be involved in all stages of the testing program, including its planning and the interpretation of results.

. Teachers must be prepared to deal with the results of the testing program. Briefings should be held in advance of the release of test results, perhaps using simulated data. Open discussion during faculty meetings may serve as a useful aid in preparation for discussing the actual results.

Over a dozen suggestions have been made relative to teachers' negative attitudes and responses to testing. Before such suggestions can be implemented, however, another question must be answered: where does the responsibility lie for implementing these suggestions? Is the school district responsible? The test publisher? The university professor?

It may sound overly simplistic to say that the responsibility should be shared; but if we really believe that testing, or more precisely, the information produced as a result of a balanced testing program, is an integral part of the instructional cycle and therefore critical to education, then we are all responsible.

PART III - TESTING AND THE TEACHER: OPPORTUNITIES AND OPTIONS

A STRATEGY FOR UNITING TESTING AND THE TEACHER

Roy H. Forbes
Director, National Assessment of
Educational Progress
Education Commission of the States

There are two primary purposes for a testing program within the context of the needs of a local school system. The first purpose is to aid the teacher in the diagnosis of learning difficulties and abilities that individual students may have. This area includes the measurement of individual student progress. The second purpose is to assess the system's instructional strengths and weaknesses. This area includes needs assessments and group performance measures. Perhaps someday testing technology will allow us to serve both purposes efficiently and effectively with the same instruments and procedures. But we do not have that technology today. When present systems attempt to do both, they prove to be inefficient--in terms of fiscal resources and time of teachers and students.

A cost effective and a time efficient testing program must take into consideration the three major instructional roles played by the teacher: diagnostician, developer, and implementer. The teacher is the key diagnostician in the implementation of any instructional program because the teacher has the responsibility of making day-by-day instructional decisions. Some of these decisions are based on group information. For example, following instruction on the algorithm for computing percentages,

a teacher is interested in how many in the group can successfully apply the algorithm. This provides a measurement of the overall effectiveness of an instructional strategy. Also of interest is the identification of common mistakes made in trying to apply the algorithm. This provides the teacher with information needed to refine the instructional approach used in teaching the algorithm.

The teacher also needs to make decisions using information about individual students. For example, did an individual uniquely misapply the algorithm; was one student's performance different from that of other students? Answers to these questions inform the teacher how to tailor instruction to the individual.

To use test information effectively, a teacher must know the context and limitations of the information. For example, does the test item match the instruction? There is no need to test the application of an algorithm if the algorithm was not taught. Also, there is no need to test the first-level application of a concept within a context that is "foreign" to the student. Only after the concept has been grasped in a familiar setting should the testing attempt to measure application of the concept within a context which requires additional skills beyond the simple application of the concept. To illustrate--the application of the percentage algorithm should be initially measured with computational items and with "word problems" in a context familiar to the students. Problems should not be in a context with which students are not familiar, that is, asking 13-year-olds in Florida a question pertaining to the percentages relationship between snow depth and moisture.

The teacher needs to know the instructional material, the instructional process, the academic and cultural backgrounds of the students, and the group dynamics of the particular classroom. The teacher also needs to have the ability to use this information in selecting or creating a test most appropriate for the situation. Assistance can and should be given to the teacher in the form of complete tests and items to select from or use as guides, in-service training and support in the development of skills, equipment and materials required to produce teacher-made tests.

The teacher has available three types of data to help with the role of diagnostician. First, and most important, is the information from teacher-made tests. The term "teacher-made test" is used in its broadest sense. A test need not be an objective, multiple choice, paper and pencil instrument. A writing essay, an oral report by an individual or a group, a project requiring the application of a scientific principle--all of these should be part of the teacher's repertoire of assessment methods.

Even though a strong case can be made for the teacher as the nucleus of a diagnostic testing program, it is unrealistic to expect each teacher to be trained in the application of sophisticated individual testing instruments and procedures. Each teacher should have available the services of a diagnostic testing center staffed by trained measurement specialists. This would provide the second type of data available to the teacher as diagnostician. When a teacher has reason to believe that additional individualized testing would be beneficial in identifying

either learning disabilities or abilities of a student, then these services should be used. The teacher would participate in the testing process, if appropriate, but responsibility for the service would be assigned to a trained test and measurement specialist. Results of the testing should be discussed with the teacher and, when appropriate, with others responsible for a student's education--parents, the student, the principal, the instructional team leader.

At present,this type of service is available through special education programs. What is suggested is expanding the services to be available to all students. It is not anticipated that all would need the service but service should not be restricted. Of special interest would be the use of the diagnostic testing service to identify the gifted. Sometimes the gifted student who comes from an economically disadvantaged background is overlooked. The procedures and instruments used to identify the gifted should be selected to identify these students as well as those from more economically advantaged groups.

The teacher should not be the only person to use the services of the diagnostic testing center; parents, principals, instructional team leaders and individual students should also have access to these services. The "teacher-as-the-nucleus" and the diagnostic testing service would provide the part of the testing program designed to be responsive to the instructional needs of individual students, the primary purpose of a testing program.

The third type of data available to the teacher as diagnostician generates from the second purpose of a district's testing program: the

assessment of the system's instructional strengths and weaknesses. A system-wide assessment program can be a cost and time efficient method of determining how well a system is accomplishing its goals. It is not necessary to administer the same test to each student in a given grade. Also, it is not necessary to rank individual students; group scores are sufficient for system level information. A matrix sampling process is most cost efficient. This allows for the collection of a broad range of information from a sample of students and cuts costs as well as student time involved in system-wide testing.

Although the primary purpose of this segment of the testing program is to provide system level data, it does provide information that a teacher can find helpful. What areas of the curriculum need to be stressed for all students or for some students in the classroom? What are the most common errors made by students in responsing to different types of questions? Answers to these and similar questions can help the individual teacher to become a better diagnostician.

Diagnostic data from these three sources, teacher-made tests, individual diagnostic testing services, and system assessment programs, can provide guidance to the developer or selector of instructional strategies whether it is the central office staff or the classroom teacher.

One example of the use of testing data and materials to assist in developing an instructional strategy is provided by the National Assessment. Initially, the Assessment used a holistic approach in scoring writing, but this was not considered to be very satisfactory in measuring whether an

essay accomplished its intended purposes. Through work with consultants, staff developed a different approach to scoring writing. The primary trait method is used to measure the extent to which an essay satisfies its intended purpose. CEMREL sponsored an effort to extend the primary trait assessment technique to instruction by developing a handbook designed to assist elementary teachers in the instruction of writing. The handbook and other descriptive information about the primary trait technique can be used by both system-wide curriculum developers and individual teachers as they plan instructional strategies.

Another example of how assessment results can be used to assist the teacher as a developer is the series of articles which appeared in The Arithmetic Teacher following the first National Assessment of mathematics. These articles, using information about the types of errors students had made, discussed instructional techniques that the teacher could use to help reduce the likelihood of these types of errors.

Curriculum developers create materials for the teacher. But the teacher using diagnostic information must adapt these materials for use in the individual classroom. In this sense the teacher must also be a developer.

When needs have been diagnosed and strategies developed to respond to the needs, then the teacher becomes the implementer. Objective or criterion-referenced items provide an effective tool for the teacher to use in determining whether the instructional strategy is resulting in the desired outcomes. Teacher-based formative evaluation should be a part of any instructional program. This activity loops back to the

teacher's role as diagnostician.

Testing materials can be used in helping the teacher with this step. The most desirable system would provide the teacher with the opportunity to quickly scan items that could be used to check student progress. A computer terminal with a cathode ray tube would provide an efficient means by which a teacher could quickly review and select items appropriate to use with a specific lesson plan. Item selection should follow lesson planning, not the other way around. This type of system would be relatively inexpensive compared with the time required for each teacher to develop unique items and would be vastly superior to a system where items are pre-selected by individuals other than the teacher. Once data are collected the teacher shifts back into the role of diagnostician.

All of the teacher's roles are equally important. Test information is useful in each role. But to make any instructional system work, the key element is the teacher as the diagnostician, as the developer and as the implementer. Competencies in these areas can be developed and improved through in-service training and instructional support systems. Tests should not be used as a teacher accountability mechanism. What should be expected is a well implemented instructional program with testing as an integral part of that program.

MAKING TESTS MORE MEANINGFUL FOR THE TEACHER

Wilson Riles
California State Superintendent
of Public Instruction

In my view testing and teaching are natural partners in the education process. Therefore, I welcome this opportunity to express my ideas because I do have some strong opinions in this area--opinions honed by some forty years of attempting to make schooling more effective.

Testing is certainly not new. The Chinese are credited with inventing it five thousand years ago to assess the merits of those who wanted to be part of their civil service. The forerunners of our modern achievement tests were developed in the early 1900's and we have used them, as the Chinese did, as sorting, classifying, and screening devices. It is only in the last three decades that tests have been used extensively in the schools. For an art form to be around this long--five millenia or even five decades--it disturbs me that more use has not been made of the information potential contained in the results from assessment instruments. If teachers take a narrow view of testing, seeing the process simply as a reporting device, they are overlooking some significant opportunities. Let me illustrate this point.

In California, we have embarked on a series of educational reforms. The keystone of these efforts is assessing and understanding student needs and organizing and implementing an effective program to meet them. An appropriate assessment program can be a major factor

in understanding the student. Yet, many assessment programs are viewed as devices used by the district office for the state or federal government, whose sole purpose is to provide information to somebody else, not the teacher in the classroom. There are frequent cries about too much testing, too much paperwork. In many instances, the programs are imposed on teachers with no clear sense of what use can and should be made of the test results. Indeed in many cases, such as Title I evaluations, teachers never even see the test results. It is a small wonder so many are disaffected, that so many are blind to the opportunities inherent in assessment to understand a student's needs and to organize and implement an effective program to meet those needs.

I do not want to be too pessimistic, however. I believe that we--principals, evaluators, and educational writers--can improve our practices. We can train teachers to realize the potential value of the information gathered in sensible assessment programs. The role of the teacher is to plan, to organize, to implement, and to judge individual progress. It is a complex and demanding job and teachers need all the help we can give them. I might add that learning is a complex and demanding job too, and students need feedback on what they are doing right and what and why they are failing to learn.

The so-called help we often give teachers--and I have some awesome illustrations--are prime examples of what not to give them. What not to give them is complex material full of technical jargon and mysterious statistical symbols. What not to give them is an additional clerical burden. What not to give them is a useless tool to clutter their already

overloaded "bag of tricks."

What we should give them is obvious: clear, unambiguous, immediately useful materials, which can be used in the day-to-day instructional program, materials which can be used effectively in the feedback process the student so desperately needs to maintain forward momentum.

I have deliberately interchanged the word "assessment" with the word "testing." When I say "test" many people immediately assume that I am talking about the standardized norm-referenced test. It is clear to me that a functional assessment goes far beyond that type of instrument. There should be many options available to teachers if they are to realize the opportunities of testing--options which will give teachers assessment information to assist in the total instructional program. Among these options will be the criterion-referenced test which, in many instances, has more utility than does the standardized test. But I think we should open up other options for teachers to use in the classroom as well, to help them gather information from a variety of sources which they can integrate into their programs along with conventional testing data. We have to look beyond the conventional, the formal, the statistically-oriented approach and find new ways of organizing the kinds of information teachers have always collected about students through observation and work samples.

The more information we have about the objectives we are trying to achieve in our educational programs, the more effective we can be in assessing the effectiveness of our efforts with each particular child;

the more effective we are in assessment of our own progress, the more helpful we can be in ensuring the student's progress. Tests are nothing more than a formal way of organizing and indicating levels of knowledge. Many times, less formal, less systematic, but more valuable insights can be gained by increasing the sensitivity of the teacher to what goes on, not just in the classroom, but in the mind of each child, as well. I cannot stress the point too strongly that I believe that assessment for assessment's sake is a waste of effort. Assessment for the child's benefit is the ultimate reason for any testing program.

As I am speaking with you, there are 38 states which are embarking on a new endeavor which is popularly called "competency testing." California will also embark on this journey. I can assure you that we have learned much from the Florida experience. What disturbs me is the popular conception that these tests will be a new hurdle on the way to acquiring a high school diploma--assessment for the sake of assessment, sorting the worthy from the unworthy. The media leaps upon the horror stories of how many students will not get a diploma. That is newsworthy; that is one more spike driven into the coffin the doomsayers are building--the coffin prepared for the burial of the public school system.

What does not make the evening news, what does not make headlines in the press, is the intent of what we are doing. That intent is the early identification of student deficiencies, in our case at the fourth grade level or even before. Through this early identification and intervention, we are using the competency or testing movement to focus

on improvement of basic skills in our population. I hope, by the way, that this focus on basic skill acquisition will not be to the detriment of a well-rounded program. I am sure with intelligent implementation it need not be. I look upon the competency testing movement as a welcome mechanism by which the total educational program can be made more effective for any child.

I see in assessment the options and the opportunities basic to teacher effectiveness. And teacher effectiveness is the rock upon which schools should be built.

ISSUES IN TESTING: THE PERSPECTIVE OF ORGANIZED TEACHERS AND PROFESSORS

James G. Ward
Director of Research
American Federation of Teachers

One of the foundations of Western civilization is an underlying rationalism and faith in the scientific method, manifested in Western thought from the philosophies of the ancient Greeks through the Renaissance, the Enlightenment, the Pragmatic movement in late 19th century American philosophy to the unbounding faith in futuristic technology and scientific innovation we embrace today. This faith in systematic observation, classification, explanation, and prediction has moved man from pastoral innocence to the wonders and achievements of the modern world.

Education has been not only an integral element in the advancement of Western culture, but also the beneficiary of the very process it carried forth. Over the past eight decades, American education has embraced the rational moral philosophy of Dewey, survived the rampant behavioralism of endless psychologists and sociologists of education, and withstood the recent debates of humanists and technocrats over deschooling, equality, opportunity, values, futurism, and the like. A critical element common to all these phases has been the centrality of the rational or scientific approach to studying and predicting human behavior.

To anyone even remotely involved with the educational enterprise,

it is evident that one part of this rational approach is the derivation of information about students, programs, and processes for the purpose of making decisions about them. Decisions need to be made about student placement, student accomplishments, student needs, program successes and deficiencies, and the value and weaknesses of particular processes in education. As in any undertaking, decisions must be made even though bases on which decisions are made are not an exact science. However, the critical question is on what bases they should be made and with what information. Testing is a general construct which purports to provide some of that information for making such judgments, but such information, used alone, has extremely limited use.

Testing is now under attack. Consumer advocate Ralph Nader draws allusions from his earlier study of production of human death wrought by General Motors' Corvair when speaking of the effect of testing on children. And since 1972, the National Education Association (NEA) has called for a moratorium on all standardized, norm-referenced testing. Its close allies have followed suit with equal rancor and vigor. For example, the national PTA has attacked current testing practices with uncharacteristic vehemence; the National Association of Elementary School Principals has published a book on the subject which is an openly advertised one-sided attack on testing. Further, in a time when the public has been deriding government regulation in general, New York and California have adopted legislation regulating test producers and the use of tests, and a number of other states are considering similar legislation.

Is all of this long needed reform or a Luddite-like attack against modern educational technology? Is this an insidious attack on the educational establishment or a cleansing of Aegean stables of the American public school? Most importantly, is the attack warranted or is it just something which will pass with time? This paper cannot provide a definitive answer to all these questions, but it can and does convey the views of elementary, secondary, college, and university faculty on the subject of educational testing.

As we all know, many people attack tests because of the messages the tests have been providing. Test results have indicated that children are not learning the basics, that literacy levels are not what we expect them to be, that college entrants are not as well prepared academically as they once were, that college graduates of today are inferior in their competence to the graduates of yesterday, that blacks and other minorities in some parts of the nation are receiving substandard educations, and that our society is in some kind of general academic decline. A response to these conclusions drawn from testing results, a response adopted in one form or another by most testing critics, has been that there is really nothing wrong with children of today or with our society but rather that tests are at fault for three reasons: (1) tests do not measure what is really important; they only measure a very small area of what education should provide; (2) tests do not accurately predict success in school, college, graduate school, or later life; (3) tests are biased against the poor and minorities. The crux of the argument is that tests just do not do what they are supposed to do.

Barbara Lerner, an attorney, psychologist, and a specialist in testing and learning, has said we really have three options to the bad news brought by tests: (1) we can respond to this bad news that tests bring us by attacking them and trying to destroy public confidence in them. This is essentially what the NEA, Ralph Nader, and the other critics of tests have been trying to do; (2) we can recognize the very real problems that show up through test results, but try to deny responsibility for them. We can blame incompetent teachers, indulgent parents, or unworkable schools instead. Some of the "new left" education critics of the late 1960's tried this approach; or (3) we can squarely confront the truth brought by test results and attempt to do something constructive about it. This includes improving the technology of testing as well as improving the educational process itself. The American Federation of Teachers (AFT) has adopted the last approach and believes there is strong empirical evidence as well as logic to support such a choice.

Some Empirical Evidence

In 1979 the AFT received a grant from the National Institute of Education, Department of Health, Education, and Welfare, to develop materials and provide training on teachers' use of standardized tests. As part of that project, the AFT Department of Research conducted a survey of teachers who were chosen to be representative of the AFT membership at large. The survey results provided some valuable information about the attitude of teachers as the prime users of test data.

Teachers in the survey, for example, felt that standardized test results were most useful for individual student placement and the diagnosis of student needs. On the other hand, they found test results least useful for modifying course content and evaluating instructional programs. When asked what improvements in tests or in reporting test results could be made to make tests more useful, teachers indicated the following:

- Results should provide more prescriptive information, such as guidance as to what materials, instructional activities, or programs are needed.
- Results should provide an adequate profile of student strengths and weaknesses.
- Results should be returned in a timely fashion for use by the teacher.

In short, teachers believe that tests are useful, or could be mainly useful, for diagnosis and prescription.

In the survey, teachers were also asked what information would be derived from a "perfect test," if one were constructed. Far and above any other elements, teachers wanted tests which provided prescriptive information for each pupil and which provided a profile of students' strengths and weaknesses. Again, teachers want information for diagnosis and prescription. Likewise, in a 1979 study of teacher attitudes toward tests conducted by Michael D. Beck and Frank P. Stetz of the Psychological Corporation, the most frequently indicated use of tests by teachers was for "diagnosing strengths and weaknesses."

Testing Regulation

A related issue in the testing field which has recently been in the center of public debate has been testing regulation legislation or what is more commonly called "truth-in-testing" legislation. Testing regulation proposes that the producers of tests used for student selection into post-secondary educational institutions make public a vast array of information about those tests, including the test questions. Its proponents claim this legislation is an attempt to "open up" the testing industry and cast off the shroud of secrecy which surrounds it.

Albert Shanker, AFT president, in his testimony before the Subcommittee on Elementary, Secondary, and Vocational Education of the U.S. House Committee on Education and Labor, posed two questions for consideration: (1) what are the pressures and problems testing regulation laws attempt to address; and (2) do they provide appropriate responses to those questions? The AFT believes that the lack of any regulation in an industry or enterprise invites misuse, abuse, and error in the processes of that undertaking. Error in test score reporting on selection tests, such as the Scholastic Aptitude Test (SAT) and others, have been documented, and misuses and abuses of such tests, including overreliance on their results for selection purposes, have been reported. For these reasons, the AFT has always supported the concept of testing regulation or truth-in-testing.

One of the reasons for this posture is firmly imbedded in the historic position of the American trade union movement in support of open access to post-secondary education for all students who can benefit from the opportunities provided by such experiences. Selection tests, such as the

SAT, when used in conjuction with other data, such as high school grades, provide a good indicator of later success in college. Given the wide variety of post-secondary institutions available with their vast array of programs, coupled with the high cost of higher education in time and money to the individual and society, placement in the appropriate institution and program is a decision of momentous importance. Selection tests do provide much valued information to make placement decisions efficiently and effectively. Some detractors of selection tests need to be reminded that such tests never prevented students from receiving a post-secondary education but have helped place millions of students in the proper college or proper program.

Teachers and professors agree with the premise that the testing regulation seems designed to ensure the proper and fair use of certain standardized tests in selection and placement of students in appropriate post-secondary educational institutions. They see selection tests as diagnostic and not discriminatory and, therefore, regard the concept of regulation of such tests as necessary.

As to assessing the appropriate responses, this they find to be more difficult. The Educational Testing Service (ETS), and the other test producers, claim that the "truth-in-testing" law adopted in New York State, as well as two similar bills introduced in 1979 in the U.S. House of Representatives, go too far by mandating the disclosure of too much information. The proponents of the legislation, on the other hand, stand behind their contention that "full disclosure" is needed. The debate is not enhanced by the posture of some of the regulation proponents who are

self-proclaimed Davids attempting to slay the Goliath of the testing industry with its gargantuan powers and influence. Nor is any light added by ETS and the Ralph Nader groups who each proclaim the superiority of their statistics without providing clear and cogent support of their positions.

The position of teachers and professors is quite clear. These tests are very useful for identifying strengths and weaknesses a student brings to post-secondary education. When used with other information, they provide data which allow a student to make best use of his or her time and opportunity. Enough disclosure of the processes of test development, the test instruments themselves, and test scoring and reporting needs to be ensured to guarantee fairness and eliminate abuse and misuse. Quality is critical to the process. However, teachers and professors seem singularly unimpressed, uninterested, and impatient with power struggles and battles of ideologues in wars for corporate advantage. Whether ETS, Ralph Nader, the PTA, or the New York State Consumer Protection Board wins is not nearly as important to teachers and professors as having fair and accurate diagnostic measures for student selection and placement. Threats exist to reaching this goal from those who want no disclosure of test content and procedures, leaving open the real possibility of error and abuse, and from those who want such massive disclosure that it would destroy test reliability, validity, and quality.

Competency Testing

In the AFT survey of teacher attitudes toward testing, about 43 percent of the respondents voiced a favorable attitude toward minimum competency

tests for students, 34 percent voiced a neutral view; and 23 percent voiced an unfavorable view. The large favorable view is somewhat surprising, given the fact that these teachers reported they had little involvement in the development of the tests, had little to say in setting passing scores, and believed the tests did not reflect classroom objectives or provide information for planned remediation. The key to these teachers' attitudes probably rests with the proposition that teachers, like the community at large, want to know how well students are performing and the teachers, in particular, want additional information for diagnosis of student needs so that prescriptive measures can be designed and implemented.

The AFT believes that competency testing, administered early enough and well, can provide teachers with important information on student strengths and weaknesses and can allow corrective measures to be taken where weaknesses occur. Competency testing administered late or poorly will become a sham and a waste of time. Teachers, above all others, know this.

Some people in education, and increasingly many in the legal profession, are challenging minimum competency tests on civil rights ground. An interesting example is the Debra P. v. Turlington case in Florida where a judge found the state's competency testing program in violation of the rights of Florida's students and discriminatory against black students. While some have charged the judge's treatment of the equal protection clause under the 14th Amendment to the U.S. Constitution to be expansive, the court's decision delayed imposition of the sanctions

for failing the Florida tests. A central fact in the case was that more black students failed the test than white students, a disparity anticipated by Florida education officials. Common sense and history tells us that the anticipated poorer performance by black students was the result of their inferior education. Rather than taking the logical step of required improvements in the education of those students, the judge, in effect, mandated that these students be given diplomas confirming the acquisition of certain skills and knowledge when, in fact, it was shown that these students did not possess such skills and knowledge. This is a prime example of the first response Lerner described of burying the evidence, rather than dealing with the source of the problems. This does not involve civil rights; it involves civil irresponsibility.

What then do teachers want? They do not want deliberate falsification of credentials, but they do want useable diagnostic tools with which to improve the education of all children.

Testing, Teachers, and Public Responsibility

At a 1978 national conference on testing, a representative of a national parent group stated, "I come to you reflecting a deep and growing concern about the use of norm-referenced achievement tests and the virtual con job that I think school systems have done on most parents regarding the meaning of those test scores." At the same conference a teacher representative said that if teachers totally object to testing as a means of measuring performance, parents will wonder what the teachers are trying to hide, and taxpayers may refuse to support educational programs.

Decisions need to be made about students. These are judgments at best. These judgments include in what program the student should be placed, to what college the student should go, what kind of instruction the student should receive, who should be chosen for a particular program with a limited number of spaces, and many similar decisions.

Teachers, professors, guidance personnel, admissions officials, administrators, and others need valid information with which to make those judgments. Well-constructed tests which do what they are intended to do and which are used wisely will provid· some of that information. To do away with such tests will simply open the door wide to dependence on more subjective measures and will lead to more inequity and abuse in placement and selection. Teachers and professors are aware of this and they know that intelligent and proper use of tests is one aid in fulfilling their responsibility. They believe that tests can be improved and that we must all work diligently toward that goal. To do so, we must avoid the ample rhetoric and get down to the real task at hand--that of improving the art and science of educational testing.

PART IV - TESTING AND THE PUBLISHER: OPPORTUNITIES AND OPTIONS

TESTING: A TOOL FOR INTERVENTION

Oluf M. Davidsen
President
American College Testing

The focus of this conference--uniting testing and teaching--is one that should be obvious but often is not. The purpose of teaching is to communicate to the learner understanding and acquisition of new knowledge and skills. The purpose of testing is to determine that such skills and understanding have been attained and that the student is ready to proceed to next steps in the educational process. To separate teaching and learning from testing and evaluation is as absurd as separating diagnosis from treatment or effort from outcome.

But, unfortunately, this is often the case and disunity and disagreement prevail. Slogans and symbols have come to dominate the debate, and the central educational issue--the relationship between teaching and testing--has been ignored. Much is being said about testing and test results, vis-a-vis access, educational standards, selection, fear and anxiety, but we hear little about testing and evaluation, vis-a-vis teaching and learning--all of which makes the conference topic pertinent and timely.

My comments on the topic at hand are confined to the interaction of testing and teaching during the period of transition from high school to postsecondary education. I plan to divide my comments into three parts. First, I would like to identify some of the characteristics of

the current transition environment--to look at the facts that pertain, to examine the assumptions we make about them, and to determine that the conclusions we draw and the positions we take are relevant and sound. Second, I will point to some communication problems and opportunities within testing and educational measurement. Thirdly, I would venture some ideas for different and nontraditional approaches to measurement in the transition context.

The Transition Environment

The current debate about testing builds heavily on the question of college access or non-access. In the present age of truth in lending, advertising, and testing, I am concerned about the picture being presented to the public about the college entrance environment. If we are to hope for a more rational dialogue in the present debate, we must seek first to bring about a more realistic perception of the current admissions scene.

The typical media article on the topic invariably seems to be introduced by a description depicting white-knuckled students with sweaty palms and broken pencils, suffering from anxiety, trauma and chagrin. A doomsday atmosphere is conveyed revolving around selection and rejection with standardized tests as the dominant instrument of judgment. The inference is that many are called but few are chosen, that the competition is fierce, and that the institutions are somehow stratified into a divine order of worth.

The fact is that we are faced today not with an access problem but with a problem about how to fill the spaces available. Colleges are scrambling for students. Recent data indicate that of those who apply, 97% are admitted and 91% at the institutions of first choice.[1] The vast majority of these students are admitted on the basis of their prior school accomplishments, though in some cases, test scores do make a difference. Here I would like to note that recently I had occasion to ask some thirty college administrators this question: Does use of test results at your institution result in more rejections or admissions? Eighty-five percent responded that they led to more admissions. It should be obvious then that the admissions scene being portrayed is twenty years old and that it reflects only a small portion of the total higher education picture.

If we are truly concerned about access, I recommend that we concentrate on two other aspects: (a) how to finance the college education, and (b) what happens after admission. It is what happens in the counseling center, the academic advisor's office, in the sectioning and placement process and, especially, in the teacher's office and classroom that can turn access to an institution into access to meaningful academic pursuit and learning. It is at these points that opportunities exist for uniting student information, including test information, with teaching and learning.

Present-day Testing and Test Result Cases

Turning now to present-day college entry assessment and test result uses, what options and opportunities can be identified for uniting teaching and testing? Today, there is a renewed interest in educational measurement,

generated in part by a perceived link between testing and various social issues. To most people, psychometrics is probably one of the dullest topics around, but if it is personalized and tied to prevailing social and economic issues, the interest rises and it becomes saleable as a media topic. We have obviously reached this point, and therein lies the opportunity.

By and large, the measurement and education profession has done a poor job of communicating and presenting its case to the public. We have tended to talk among ourselves and in our own language; and in our quest for accuracy and fear of loss of exactness, we have shied away from translating that language into commonly understood terminology.

In college entrance testing, we have failed to make it clear that we are operating in a very short-range, but critical, forecasting environment. Survival or mortality generally is determined during the first college year. To aid in this critical period, we collect recent performance information and project it forward one or two years. In contrast, critics seem to assume that entrance test results are somehow everlasting, and if they are subject to change they are suspect or useless or both. This position is especially evident in the coaching controversy.

We have failed, also, to communicate the relatively high degree of accuracy of tests as tools of prediction and their unique characteristics in this respect. To me, it is quite remarkable that, in a period of three and one-half hours, it is possible to generate information equal in predictive power to classroom performance information gathered over a

period of two or more years. The predictive accuracy is, likewise, remarkably high. In three and one-half hours, information is obtained that accounts for 25% of the freshmen college grade variation. That exceeds the predictive accuracy of the typical medical diagnosis, and I am certain every dice thrower in Las Vegas would be happy to be half as accurate.

In addition, the close relationship between test results and high school grades is being cited as a reason why test results are unnecessary. In my opinion, a 12-13% average reduction in prediction error resulting from combining the two predictor variables is important, especially if one is concerned about borderline students.

Finally, I believe individual students deserve at least one opportunity to present themselves independent of others' interpretation of them (grades, recommendations, interviews). Anyone who has taught school knows well the temptation to take into account neatness, cleanliness, timeliness and prior impressions when evaluating a piece of homework or a class assignment. We have an opportunity and an obligation to do a better job of communicating the value of standardized testing and in language that can be understood by everyone. In doing so, I believe there is a particular need to deal with the purposes and utility of predictive information. It is common to view the college grade prediction as fate, as an indicator of something everlasting and unchangeable, rather than as a source and opportunity for intervention.

We are faced with basically four alternatives: (1) we can reject

test information as a predictor on the grounds that it is not adequately predictive or that it is redundant. The evidence does not support this alternative; (2) we can limit the use of predictive data to sorting and selecting purposes in which case we forego the opportunity to bring the data to bear on educational planning and the teaching-learning process; (3) we can reject the criterion (college grades) as not being relevant given the meager relationship it has to post-college achievement; and finally, (4) we can use the prediction for intervention purposes. If the prediction is favorable, we can reinforce it, and if it is unfavorable we can look for ways in which to upset and foil it through instructional intervention. In my opinion, this latter alternative presents the greatest opportunity for uniting testing and teaching toward improved learning and student development.

Before leaving the present-day measurement situation, I would mention one form of uniting testing and teaching that I fear will, but I hope will not, come about. I refer to legislation requiring the release of current test items. Because of the ingrained notion that some colleges are better than others and that higher test scores are in and by themselves better, I am concerned about the resulting pressure on schools to teach for tests. Test content should reflect the curriculum but not dictate it. We have an opportunity and an obligation to warn against this form of unification.

Nontraditional Approaches to Educational Assessment

Taken as a whole, the American postsecondary education system has come a long way in the past 25 years in recognizing the new heterogeneity

of the student population it serves--in recognizing that higher education is no longer reserved only for those who can be expected to progress most easily through the system. But, despite the diversification in educational opportunities and instructional approaches (community colleges, evening and Saturday classes, home study and educational TV, etc.), we continue to operate on the basis of a relatively limited and narrow set of standards and measures for assessing incoming students' competencies. Our assessment is very much geared to the college curriculum, the traditional instructional mode and prevailing evaluation methodology. This, I repeat, is very legitimate and necessary. The consequences of abolishing the use of test results and high school grades at the admission point is simply to invite a higher dropout rate. The addition of interviews and letters of recommendation to the admissions data base will do little to broaden the assessment because they inevitably reflect the present perspective about college education and what it takes to "succeed."

In an attempt to broaden student information in a different manner, ACT began some fifteen years ago collecting and reporting information to colleges about students' actual accomplishments in out-of-class activities, including science, leadership, art and music, community affairs, and writing. Examples of questions from these nonacademic achievement scales are:

Have you won a prize in a regional or state science fair?

Have you been elected to one or more student offices?

Have you participated in a state music contest?

Have you worked as a volunteer on a civic improvement project?

In doing this, we find that the average student responds affirmatively to some 12-13 questions out of the total of 63. We also find no difference in response rate between minority and majority students.

Further, we have been interested in looking into the relationship between three sets of factors:[2] (1) pre-college student information in the form of high school grades, standardized test scores, and the non-academic achievements; (2) college grade point average; and (3) post-college achievements as measured by adult nonacademic achievement scales similar in content to those used at the pre-college level. While these adult achievement scales are by no means the only proper outcome measures for a college education, they seem to represent some of the cultural and citizenship goals a liberal arts college may well strive for. The relationships between these three sets of variables are quite interesting and perhaps revealing:

1. The correlation between high school grades and test scores is quite high ($r = .51$) as is the relationship between these factors and college GPA ($r = .34$).
2. Pre-college out-of-class accomplishments are not related in a significant manner to high school grades or test scores.
3. High school grades, test scores and college GPA are not significantly related to adult achievements as measured by the adult achievement scales ($r = .05$-$.09$), nor for that matter to most any other measure of adult achievement.[3]

4. Pre-college nonacademic accomplishments are, however, related to respective adult achievements at about the same level of magnitude that high school grades and test scores are related to college GPA (r = .26 for graduates and nongraduates combined).

Without underestimating all of the shortcomings of this data, it suggests that there is promise in searching for additional indicators of potential beyond those revealed in test scores and grades.

The use by colleges of the nonacademic achievement information is limited to general educational advising and for guiding and encouraging students to continue these interests and pursuits through campus facilities. Even so, I believe there is promise in continuing the search for additional indicators of potential beyond those revealed in test scores and grades, and I would much rather be able to use our resources for this purpose than for tripling our test construction output as test legislation would require.

One word of caution, however. There is much concern and suspicion today about bias in standardized tests, coupled with accusations that they do not measure other important dimensions of human potential, e.g., motivation, stamina, resiliency. We must be aware that measurement of these noncognitive traits, whether motivation, learning style, or pertinent prior experience involves a much higher risk of reflecting cultural bias than does construction of standardized tests of educational development that have their base in the high school and college curriculum.

A primary reason for the absence of a demonstrated relationship between in-college and subsequent achievement is, obviously, the difficulty in identifying a proper criterion measure. Another possibility is that the total impact or gestalt of the college experience may elude us. It is possible that the sum total of all college grades earned, the GPA, is the sum of largely unrelated elements that fail to portray the intellectual stimulation, maturation, and learning that has occurred. In our society in particular, we become impatient when we cannot see clearly the relationship between effort and outcome. We think it makes sense to identify desired outcomes and then to find and apply means and techniques for achieving them.

In reference to this point, I would briefly mention two efforts currently in progress, one at UCLA and the other at ACT. Using what they call Quality of Effort indicators derived from a questionnaire, Robert Pace and Jack Friedlander[4] obtain information about how frequently students use college facilities--libraries, conferences with teachers, theaters, science laboratories, etc. When they relate these indicators to self-assessed student attainment variables, they find that the Quality of Effort indicators add significantly (10-15%) to the explanation of variance in attainment. Even though their predictor and criterion measures seem to be somewhat inbred, there is a suggestion here that we are overlooking an important intervening variable in explaining college outcome, namely the amount and kind of effort students put into the college experience. Perhaps the way you go to college makes a difference.

I have suggested that, in and by itself, the college transcript may

be alphabet soup without much discernible flavor or nutritional content. Based partly on this possibility, we have been involved at ACT, for the past four years in cooperation, with scores of college and university teachers, in the development of general college education outcome measures. The criteria or outcomes projected are in six major areas: communicating, problem solving, value clarification, functioning within social institutions, using science and technology, and art. The measurement devices are of three types: (1) A series of fifteen simulation activities are based on stimulus materials drawn from the adult domain--activities that require application of knowledge and skills to issues and problems commonly encountered in adult life. These responses are judged by faculty members using standardized rating scales. (2) A parallel component uses similar stimulus material but relies on multiple-choice responses. (3) And another component is in the form of an activity inventory which inquires about the extent and level of actual participation in various activities in each of the six outcome areas.

It is my belief that this type of evaluation--call it testing if you will--can help build a bridge between what goes on in the Political Science 104 classroom and those purposes for which Political Science 104 is part of the university offering, a bridge that will eventually lead to the broader educational objectives students had in mind when they enrolled in the university and signed up for Political Science 104.

REFERENCES

[1]Tuition and Student Aid: Their Relation to College Enrollment Decisions. Washington, D.C., American Council on Education, 1978.

[2]"Varieties of Accomplishments After College: Perspectives on the Meaning of Academic Talent." L. A. Munday and J. C. Davis, ACT Research Report #62, Iowa City, Iowa, 1974.

[3]"The Relationship Between College Grades and Adult Achievement: A Review of the Literature." D. P. Hoyt, ACT Research Report #7, Iowa City, Iowa, 1965.

[4]"The importance of quality of effort in predicting student attainment." Jack Friedlander, Laboratory for Research on Higher Education, UCLA. Paper presented at the 1980 Annual Meeting of the Association for the Study of Higher Education, March 4-5, Washington, D.C.

TESTING AND THE PUBLISHER: CAN WE TRUST THE TESTS?

Lawrence W. Feinberg
Education Reporter
Washington Post

Shortly after Maryland gave its first round of functional reading competency tests last year, officials found something had gone seriously awry. In Baltimore, the largest school system in the state, study guides had been distributed before the testing which included two of the actual questions on the test plus several others that were very close. Of course, the results for Baltimore were invalidated--though they were counted in the rest of the state. But even a year later, the chain of problems started by the study guides continues to bedevil the tests.

Because security had been breached, the state education department rushed development of a new version of the competency test. After it was given last fall, officials discovered another serious problem: at least seven of its questions were too ambiguous to be fair. The results of this test were invalidated throughout the state.

Next year the state department of education plans to use questions from the first test again plus new ones it is developing. But clearly, no matter how good this revised exam turns out to be, confidence in the test has been seriously weakened.

As I see it, there are three quite separate issues involved in the question of how good the tests are that school systems and college

admissions organizations give millions of students each year: (1) the quality of the tests themselves; (2) the quality and integrity of the way in which the tests are given; and (3) the quality and propriety of the ways in which the tests are interpreted and used.

It is on the first point--the quality of the tests--that much of the controversy about testing has centered for the past decade. The critics, led recently by Ralph Nader, charge the tests with being biased and unfair, of not measuring academic achievement properly, and of not predicting future achievement well.

The defenders of the tests, including, of course, their publishers, but most reputable educational psychologists as well, have tried to refute these charges with item analyses, validity studies, prediction formulas, and the full array of the psychometrician's art. The evidence they have marshalled is impressive. So far, at any rate, it has persuaded me that the major national tests--the SAT, the ACT, the CTBS and the ITBS--are indeed fair and offer independent ways of measuring academic achievement. But this is not so true of the new competency tests that state school boards, such as Maryland's, are developing. These new tests may well tie in to local curriculums. But whether they are prepared with the same safeguards and rigor that the national tests employ is open to considerable doubt.

Whatever their merits, the arguments about testing usually stop here. Unfortunately, they should not. For as more and more states and school systems use more tests to do more crucial things, the possibilities increase greatly that the tests will not be properly given

and their results will not be properly used.

Such problems generally do not arise with the college entrance examinations. They are given under the tight supervision of the testing organizations. That is not the case, however, with the standardized achievement tests, published by private companies, or the competency tests, mandated by state boards of education. These exams are given by local school boards and teachers--with test manuals, of course, but with no check by the test publisher or any other outsider to see that all the rules are scrupulously followed. Alas, there must be serious doubts about the way, in fact, the tests are administered.

One description of the problem was given in a memo several years ago by a senior official of a national education research group. He asked me not to use his name, but he wrote as follows:

> Although it is customary in public education to allow schools and school systems to evaluate themselves, this is obviously unreliable and unsatisfactory. Most teachers and administrators try to administer standardized tests honestly to their pupils. But without any auditing procedure, the temptations are very great (to do otherwise). The greater the pressure for results--and the pressure is increasing with the current trend toward greater accountability--the less reliable self-evaluation becomes.
>
> The existence of "irregularities" with respect to achievement testing is common knowledge among school people....Most take the form of excessive coaching for the test, but there are more flagrant types of misbehavior as well. Some of these are teaching the particular words to appear on the test, practicing the test itself, changing the answers before the tests are scored, giving pupils aid during the test, allowing additional time, and failing to test selected pupils who are expected to do poorly. I saw or heard reliable reports of all these irregularities during school visits in seven large cities this past spring.

More recently, in Pocomoke City, Maryland, a small town on the Eastern Shore, a school principal was forced to resign after his superiors found out that students were given mimeographed copies of a standardized test to help them prepare for the exam. (Washington Post, January 28, 1978.)

In Washington, when I sought to explain why an elementary school in a poor neighborhood had high scores, I found out it had been testing in a particular grade only those students who already had shown they had achieved grade-level skills. The others were held back in lower grades and tested there. However, all children when they reached 13-1/2 years old were promoted to junior high school regardless of achievement, including some who were skipped from fourth or fifth grade. (Washington Post, June 14, 1979.)

Until ten or fifteen years ago, test scores were a private matter between students, schools, and colleges, and the pressures did not exist for officials to make the scores appear better than they were. The scores were used for placing students in different groups or classes, for advising them to take particular courses, or attend particular colleges. There was every reason to give the tests properly so that the information for making these decisions would be sound. There was no reason to help students improperly because the line between test-givers and test-takers was clear.

This is no longer so, partly I confess, because of newspapers like my own and hundreds across the country. We publish the test scores--school by school, grade by grade, district by district. We make the uncomfortable comparisons. We name the names: the sixth graders at Reed

Elementary School in Washington are about a year below those at Adams School just two blocks away. The scores in Prince George's County, Maryland, are far below those in Montgomery. Scores in Fairfax County, Virginia, are higher than in Arlington.

Communities react to the stories--with pride when the scores are up, with defensiveness or determination when they are down. Newcomers to the area call me up and ask for the test score stories, so they can make their own comparisons to help them decide in which district they wish to live. Of course, that is a treacherous business. But the scores are the only objective measure the public has of the school's academic standing. In a real sense, then, teachers and schools now are being judged by the tests along with the performance of individual students. As a result, school personnel are under great pressure--from shame, praise or perhaps promotions--to make sure students do well.

How honest are the test-givers? How widespread are the "irregularities"? We do not know. But clearly the incentives for wrongdoing are strong enough and the reports of improprieties are frequent enough that some system of monitoring by outsiders must be established. One obvious parallel is the system by which income taxes are collected. People fill out their own tax forms and figure their own taxes, but the Internal Revenue Service has plenty of auditors to make sure it is all done right.

Who might the test auditors or the outside examiners be? Of course, they might be state or federal agencies. Certainly, for any test that is mandated by a state there is a clear state responsibility to make sure it is given properly. The test publishers might also be the auditors, or

examiners, making spot-checks in different schools to see that the tests are given properly or, more radically, conducting the tests themselves with a cadre of examiners whom the publishers, not the school systems, employ.

Yes, this would be difficult. Any system of state-run testing would collide with the traditions of local control. For publishers to actually administer the tests or check too intrusively into how they are given would create serious commercial problems. The test publishers are contractors and must be chosen by the very school systems whose success they evaluate. It is a competitive business. It would be difficult for one company to be more of a stern policeman than another, partly because that might put off potential customers, partly because it might cause its prices to rise.

Yet, in other fields, systems of checking by outsiders have been established. Businesses and governments hire independent auditors, television networks hire rating services, politicians hire pollsters. The problem of getting honest information about the level of student performance can be solved. And it has to be, because unless there is a clear guarantee of the integrity of how tests are given, testing itself falls under a cloud.

There is another area of serious concern--how the test results are interpreted and used. Several years ago, I encountered two examples of this type of problem. One was in Gary, Indiana, where test scores were used as the basis for paying a private contractor to operate the Banneker

Elementary School. The superintendent, Gordon McAndrews, strongly backed the experiment, and after the first year put out a press release hailing its results. The release was accurate, but not entirely clear. It blurred the distinction between average scores and average gains. Reporters who did not know much about test scores used the release as the basis for glowing stories that spread across the country. Fortunately, I did not write one of those stories. A year later it became clear that the company-run school was not working well, and the project was dropped.

In the Washington public schools there was a similar situation where the superintendent, Hugh Scott, announced with considerable fanfare that test scores had risen sharply one spring compared to the previous fall. There was no fanfare, however, when the test scores the next fall turned out to be as low as they had been a year earlier. What happened? Had the schools suddenly improved and just as suddenly turned sour again? Not at all. In retrospect it became clear that it was all a matter of how standardized tests are normed--with level month to month gains throughout the school year and no drop over the summer.

This simply does not follow the pattern by which many children learn--particularly those from low-achieving areas who generally do not read much over the summer. Actually, the even intervals in the norming scale are no problem if the tests are given in the same month each year, which is the way they traditionally have been administered. But when they are given twice in the same year, a serious distortion sets in if anyone wants to use the results to show "normal" progress or better.

The test publishers, of course, understood all this. Whether the

District of Columbia school officials did is uncertain. However, the test publishers are employed by the school systems, and they report all their data to the people who hire them. The school systems, in turn, decide what to release to the public and how to display and interpret it.

In a decade of reporting on test scores, I have found no false data presented, but sometimes crucial things that were contrary to the interests of a school system have not been said or were not said clearly. What is the solution to this? Perhaps the test results could be reported publicly by the publishers--after giving the local school systems some time to check them over and analyze them first. Perhaps the reporting could be done by an independent government agency, which just has the job of evaluating schools, not of running them or seeking funds for them to operate.

Again, change in this direction would be difficult. But I believe it must take place and fairly soon. And the test publishers should play a strong role in helping it happen. It is in the interest of the people who believe the tests are sound to help establish the safeguards--outside monitoring and independent reporting--that can give the public full confidence in the integrity of the results and a clear understanding of their significance. Unless improprieties are prevented, there is a serious danger that they will compromise the whole system of testing that we now have and possibly cause its demise. That would cheer many of the critics of testing. But to those who seek academic rigor in the schools it would be a tragedy.

ACHIEVEMENT TESTING AND THE PUBLISHER: OPPORTUNITIES AND OPTIONS

Ralph W. Tyler
Director Emeritus, Center for Advanced Study in Behavioral Sciences and
Senior Consultant for SRA

"Until the late 1960's criticism of educational testing was largely confined to academic debate. Since then, testing has become a center of public controversy." This statement is the beginning of the report of a Conference on Research on Testing held in August, 1978, sponsored by the National Institute of Education. The preface of this report continues, "minority groups object that published tests are unfair to their children. Many teachers, parents and school administrators find existing tests unsatisfactory as indicators of a student's educational progress. The National Education Association has called for a moratorium on testing."

Sources of Criticism

What accounts for this mounting criticism of educational testing after nearly 60 years of continuous expansion? An important element is the fact that modern testing theory and practice developed when education was rationed and there was a great demand for objective methods of sorting and selecting students. Since the end of World War II, this situation has sharply shifted. Modern industrial societies find it necessary to educate all or almost all young people and the demand for means of sorting children has diminished while requests are rapidly growing for means of guiding the education of individuals and groups, of evaluation, and of

improving instruction. For these purposes, most available tests are inadequate and their data are often misleading.

They are inadequate because they do not give dependable evidence about what students have learned but instead are designed to arrange those who take the tests on a continuum from those who make the highest scores to those who make the lowest. Such an arrangement permits one to identify the position of any individual in terms of his standing in the total group; for example, Mary Moore is at the 98th percentile, and Richard Roe is in the bottom quarter of the group. Furthermore, the arrangement enables one to identify those who compose any desired fraction of the group; for example, the persons in the top quarter or those in the lowest third.

Many educators do not understand why a test, carefully designed to sort students, that is, to place them on a continuum, cannot also serve as a dependable means of appraising what students have learned. A standardized test usually includes a total set of items that can be answered by the students in an hour or so. The number of possible questions and exercises is so large that those selected for the test are a small fraction of the possibilities. Hence, to obtain a sample that is representative and adequate requires quite precise sampling procedures focused on the particular purpose of the appraisal. When the purpose is to identify the student's position on a continuum, the sampling procedure

concentrates on items that differentiate among students, largely eliminating items that may be important for teaching but do not help in sorting. On the other hand, if the test is designed to identify what students have learned and what they are not learning, the sample should be selected so as to be representative of what students are expected to learn.

The sampling procedure used in constructing the widely-used achievement tests are distorted not only by selecting items that sharply discriminate among students but also by restricting the sample still further by the requirement that an item must be one in which a larger percent of students answer correctly in each higher grade. This is in conflict with the empirical evidence that students often perform better on a topic which they have recently studied than on topics taught earlier. For example, children in the first and second grades perform better on a representative sample of items testing phonics than do those in the third grade. Phonics is usually emphasized in the first and second grades.

Another distorting factor arises from the modification of the selected items to produce smooth "growth curves" from achievement test results. School learning does not, in most cases, proceed like physical growth. The things students learn in school include definable skills, understandings, attitudes, interests, and the like. Whether students acquire these in smoothly increasing amounts over the years depends both on the particular nature of the things to be learned and the extent to which the environment of the student stimulates continuing interaction.

Because many educators do not understand the construction and composition of the widely-used tests, some common practices are inappropriate. One of these is to use an item analysis of a test for diagnosing student errors. Since the test itself provides only a small sample of student reactions, and the sample is not representative of what students are expected to learn, an item cannot be considered as representative of a class of errors. Hence, an item analysis is not helpful in diagnosis but is often misleading.

Another common misconception relates to the use of the term "grade equivalent." This is a construct which gives erroneous impressions. It assigns the "grade equivalent" to a student's score that is the average of the scores made by students of that grade. An example may illustrate why it is misleading. A sixteen-year-old boy whose score on a reading test is the same as the mean score of fifth grade students is not likely to have answered correctly the same test items as the average fifth grader. He is not usually reading the same material as a fifth grader nor making the same errors. Hence, it is confusing rather than helpful to think of him as being at the fifth grade level and futile to try to teach him as though he were in the fifth grade.

The Publisher's Dilemma

As the inappropriateness of the typical achievement tests for assessing what students are learning in school is increasingly recognized, more and more protests against their use in schools are heard. The official resolution of the National Education Association called for a moratorium

on the use of any standardized test that did not provide assistance to the teacher in instruction. Some groups are sponsoring legislation aimed to control or even to prevent the use of tests in schools. At the same time, the public concern with the accountability of schools in teaching "the basics" has stimulated legislation in more than thirty of the states to recommend and in some states to require the administration of tests to assess the skills of students in reading and mathematics. This movement is increasing the use of standardized tests.

This situation presents something of a dilemma to responsible test publishers. They view their role as providing tests and instructional materials to schools in order to help with the educational process. They try to develop materials that will meet the needs of school people and help to satisfy the demand for effective and efficient tools of learning. So-called "aptitude tests" and achievement tests were constructed and published when sorting was an important function of schools, when public education was rationed, and the accepted policy was to encourage those students who learned school tasks easily and to discourage or steer away from further schooling those students who did not easily meet the school's requirements. Both "aptitude tests" and achievement tests were useful in this sorting process. The fact that they did not help the teachers with their instructional tasks was waved aside since the sorting function was considered sufficiently important to justify the use of tests even though little or no instructional assistance was provided. Only recently have educational leaders learned that these sorting tests misled teachers, parents and the public generally about the actual

educational potential of various groups of students that public schools are now expected to educate.

But now we are no longer ignorant of these matters. What can publishers do? What are their options?

Option 1: To Continue Test Development and Marketing as at Present.

Because the general public is not aware of the limitations and inadequacies of the present achievement tests and because steps to change tests and test procedures would be viewed by some critics of the schools as efforts to evade accountability, there is a strong demand for the present achievement tests. This demand is likely to continue for several years and will only diminish gradually as there is wide public understanding of the difference between tests designed to sort students and those constructed to assess the progress of students in the instructional program of the school.

Publishers who are willing to continue to push the sales of tests of such limited value and even to invest in refinements that do not correct the inadequacies of these tests may be accused of irresponsibility. However, the issue is not as clear cut as this. Some publishers may view the situation the way some pharmaceutical houses view the sale of headache powders to persons who use them as cures for minor ailments. The producers and distributors may argue (and even believe) that a social service is rendered when the customer feels better even though the disease has not been affected by the headache powder. School boards, legislators, and other officials and leaders of public opinion seem to feel that they

are holding schools accountable when they see that achievement tests are administered and the tests reported even though the tests are related only tangentially, if at all, to the instructional tasks of the particular schools. It is possible that some test publishers feel that satisfying this desire of leaders to do something is a worthy service even though the something done has no measurable effect on the school learning of the children. In any event, there is a substantial market for the present educational tests, and this option will no doubt be chosen by a number of publishers. But it will not satisfy the critics of testing, nor help teachers with their instructional tasks.

Option 2: To Focus on the Development of Tests that Can Actually Aid the Teacher in Instruction.

Tests can be constructed that are designed to assist the teacher at several points in the process of instruction. The report of the Cape Cod Conference in August, 1978, on Teaching and Testing, distributed by the National Institute of Education, furnishes a number of illustrations of tests of this sort. For example, some aid the teacher in placing students in the instructional sequence, some identify students who have mastered the learning tasks in each unit or topic, some disclose the difficulties of those who have not, and some furnish information for parents about the progress of their children. These are a few examples of the possibilities for developing tests that can give helpful assistance to teachers in the process of instruction and in reporting to parents.

However, the refinement and standardization of tests of these kinds is not a simple matter. Just as present achievement tests are products of a development process involving the sampling of exercises that discriminate among students and the establishment of reliable samples of discriminant items, tests helpful in the instructional process will need to sample reliably what students are expected to learn in the instructional program to produce a test that can dependably indicate what students are actually learning. This is a difficult task for test publishers because their expertise has been used mainly to produce sorting tests. Hence, the staff training and the time and costs involved in the development of refined, standardized tests that can aid in the process of instruction make this option a hazardous one. Furthermore, teachers have not generally been using such tests nor have most of them understood the possibility of their development and the value they would have in aiding instruction. Hence, the demand is likely to grow slowly. And, in addition, during the current period in which the public is questioning the accountability of schools, the adoption of new tests may be viewed with the suspicion that new tests are being selected to hide declining test scores. These considerations suggest that the development costs of these new kinds of tests may not be recovered for a long time--perhaps ten years. Considering the current high interest rates, publishers may be unwilling to invest in a development process that is so costly.

Option 3: To Discontinue the Production and Marketing of Educational Tests Until the Demand for the New Kinds Will Have Grown Sufficiently to Justify Development Costs.

This third option might be chosen by publishers who do not wish to use their resources to market tests with very limited usefulness and who are unwilling to invest in developing the new kinds of tests when their short-run profitability is doubtful.

What Else Can Be Done?

Since there is a real need for tests that can assist in the instructional process, but the present demand is not great, the National Institute of Education and other sources of funding can make an important contribution by supporting research and demonstrations that could identify kinds of tests that can aid instruction and show various instructional procedures that constructively utilize such tests. The investment in development for the prototype tests used in this research and the demonstrations will only be minor because the prototypes will not be designed to be appropriate to situations generally but only to the demonstrations. The sampling problems are minimized for instruments constructed for use only in particular situations. However, the research can furnish data to guide in the construction of standardized instruments for wider use.

Effective demonstrations can help to create a more rapidly growing demand for these new kinds of tests. The demonstrations can also indicate how the tests can be used, suggest the learning systems required and show

what training teachers would find helpful in learning to use these tests.

As the research and demonstrations establish a broader base in experience and as their procedures are accepted by large numbers of teachers, publishers would then be more likely to invest resources in the development and marketing of tests to help teachers in instruction. Teacher organizations are calling for tests that can assist them in their work. Eventually, publishers will respond.

UNITING TESTING AND TEACHING: A FEW WORDS IN SUMMATION

Robert L. Ebel
Professor of Educational Psychology
Michigan State University

The participants in this conference seem to agree almost unanimously that testing and teaching ought to be united. We have considered how this might be done in the basic skills, through information services, by the teacher, and with the help of the test publisher. One note that has been sounded repeatedly is that tests ought to be used more to help teachers to teach better, less to judge them. The emphasis has been on diagnosis and remediation, and so-called formative evaluation. The emphasis has been away from using tests to assess achievement in learning.

I beg to differ. To neglect summative evaluation, the measurement of what has been learned, is to give up the strongest influence for good that testing can have on study and teaching. That we have been guilty of neglect in this area is no small part of the reason for declines in school achievement in recent decades. Testing to help pupils to learn is not an alternative to testing for the purpose of determining how much they have learned. The two purposes are strongly supplementary, if not identical, in many school situations. If tests are used to measure what pupils have learned, it will help them to learn by motivating and rewarding their efforts. It will also motivate and reward the efforts of their teachers.

Accordingly, I invite your attention to the following radical

proposition: almost all teachers should be held responsible for providing evidence on how much their pupils have learned of what was set out to be learned. It is radical in two senses. It is "marked by a considerable departure from the usual or traditional." It also goes to the "root" of the problem of excellence in education. Radical though the proposition may seem to be, it is nevertheless wholly reasonable. To show that this is so will be the purpose of this paper. The presentation will be organized as answers to four questions.

1. What should pupils learn?

Pupils learn many things in school. They learn about themselves. They learn about getting along with others. They learn physical skills. In good schools they learn moral responsibility. But the main thing they are expected to learn is useful verbal knowledge. The special excellence of human beings, among all the creatures in the world, is their ability to produce and to use verbal knowledge. It is the foundation for effective behavior, for wisdom, for moral judgment.

What is meant by knowledge in this context may be clarified by a few words on its relation to information, to understanding, and to ability. Knowledge is built out of information by thinking; it is a structure of relations among items of information. Awareness of that structure, or of a particular part of it, is what we call understanding. Seeing the rela. on of this to that, of moon to tides, of thunder to lightening, of frustration to aggression, enables us to understand these things. Finally, a mental ability is nothing more nor less than knowledge of how to do something.

In this view of learning, teacher and pupil have essential and distinctly different roles. Only the pupil can do the reflective thinking needed to transform information into knowledge. No teacher, no instructional system however elaborate, can do the pupil's thinking for him or her. Teachers do not impart knowledge, only information. The pupils must learn by their own efforts to understand the right to say , "I know." Thus no teacher should be expected to accept the responsibility for the success of every pupil in learning. That responsibility belongs in large measure to the pupils themselves.

But teachers do have a responsibility that is much more important than the providing of information. They are responsible for establishing and maintaining a good environment for learning in the classroom. Indirectly, the school administration and the community share in this responsibility, but the teacher's role is more direct and immediate. The difference between good and poor teaching is largely determined by the teacher's success, or lack of success, in maintaining a good learning environment made up of at least four essential elements: (1) a capable and enthusiastic teacher; (2) a good supply of learning materials--books, charts, films, laboratories, shops, etc.; (3) a system for recognizing and rewarding achievement in learning; and (4) a class of capable and interested learners. Few would question the first two of these. The third suggests an answer to the next question.

2. Why should learning be measured?

Learning can be fun, but often it is simply very hard work. The amount of learning acquired is likely to be roughly proportional to the

effort applied to the task of acquiring it. Further, the amount of effort applied by the learner is likely to be roughly proportional to the value of the anticipated reward for learning it. For most pupils, the immediate reward for success in learning is public recognition of that success by teachers, parents, other pupils and the community generally. The basis for that recognition is ordinarily the pupil's test scores and grades.

Some educators object to the "extrinsic" rewards for achievement that grades and honors provide. But no teacher who wants pupils to work hard for achievement in learning can afford to reject extrinsic rewards. Neither can a good school or a good community. The few educational theorists who deny the essential role of rewards in education must be quite out of touch with the realities of the classroom. These theorists have not produced, nor could they produce, evidence that pupils learn more or learn it better under the no-reward conditions they advocate.

Occasionally, of course, people do work and sometimes work very hard for rewards that are not immediate and tangible. They work to meet a personal obligation, for a sense of personal satisfaction, to achieve some ultimate personal goal. But teachers cannot rely on the intrinsic motivations of most pupils to persuade them to work hard at learning. Pupils need the more immediate, more tangible incentives that tests and grades, honors and rewards provide.

Most teachers, most schools and most societies have recognized this principle and acted in accord with it. They know that if high achievement is desired, it must be rewarded. Conversely, if the principal goal of a

teacher or a school is to support and protect pupils of lower achievement, the school is likely to produce a great many graduates who are quite satisfied with low achievement. In these days, such teachers often refer to themselves as "humanistic." But an education that refuses to recognize and to honor excellence in learning can hardly claim to produce excellent human beings.

There is another important reason for measuring pupil achievements in learning. Those measurements not only motivate and reward the efforts of pupils to learn but also motivate and reward the efforts of their teachers. They can help school officials to distinguish competent, dedicated teachers from those who are incompetent or only going through the motions of teaching.

When the learning of an individual pupil falters, it can be, and often is, the fault of the pupil. When the learning of a whole class falters, it can be, and often is, the fault of the teacher. Surely, it is true that in many of today's schools teachers do not have sufficient control over the conditions that influence learning to be held responsible for the results. But professionally trained teachers can be given, and in good schools they are given, a large measure of control over the strategies and tactics of instruction, over the classroom environment for learning. Given these things teachers may quite properly be held responsible for the general level of achievement in learning of their pupils.

Despite recent efforts to systematize or to mechanize the processes of instruction, teaching remains an art. Effective teachers, those whose pupils are most successful in learning, have a repertoire of strategies

and tactics of instruction. They select from this repertoire the ones that work best for them and are most effective with the pupils they are teaching. There is no best method of instruction for all teachers in all subjects with all pupils. Research that attempts to discover whether this method of instruction is superior to that is unlikely to be decisive or widely generalizable. The crucial question is more specific and particular. Which methods of instruction in the hands of this teacher will produce the greatest amount of learning with these pupils. The best research on teaching is likely to be action research done in the field by individual teachers.

Among the many circumstances that contribute to a good environment for learning, there is one that merits particular attention at this point. It was identified earlier in the paper as "a class of capable and interested learners." How much a particular pupil learns depends in part on the pupil's own efforts. It depends in part on the effectiveness of the teacher. But it also depends in part on the examples provided by the other pupils. If they are capable and interested they will provide models of effective study and achievement that any pupil is likely to emulate.

A number of influences have combined to place in many of today's classrooms pupils who are seriously lacking in capability for learning or interest in learning. At best, these atypical pupils distract the teacher from doing the best possible job of instruction with the capable and interested pupils. At worst, the atypical pupils interfere with the efforts of other pupils to learn and disrupt the instructional process.

If teachers are to be held responsible for, and judged by, the achievements of typical pupils in learning, they must only be asked to teach reasonably capable and reasonably interested learners. They must be allowed to exclude from their classes any pupils whose presence interferes with the progress in learning of other pupils. Special educational needs and problems should be dealt with by special teachers in special classrooms. Indiscriminate mainstreaming is a receipe for educational disaster.

3. How can learning be measured?

To reward achievement in learning, one must assess the extent of that learning. Written tests provide the most accurate and efficient means for making such assessments. Written tests are entirely adequate for assessing the extent, the depth, and the solidity of any person's structure of knowledge. That structure is made up of relations between concepts. Whether a person has a concept in his structure, and is aware of its relations to other concepts, can be determined by examining the person's responses to the questions on a written test.

Conventional wisdom among teachers is that there are many ways other than the giving of tests for assessing a pupil's educational achievements. Many of them believe that a good assessment requires the use of a wide variety of these ways. This "wisdom" is open to question. For if the focus of the assessment is on the extent to which pupils have developed structures of useful verbal knowledge, written tests can do the whole job. They are likely to do the job more efficiently and more accurately than any less systematic, less objective, less quantitative means of assessment.

Since the beginning of this century much has been learned about the art, the science, and the technology of testing. We have learned that whatever can be tested using essay tests can also be tested objectively with greater accuracy and, unless the group to be tested is very small, greater efficiency. We have learned how to write questions that test for understanding. We have learned how to test the quality of our tests. We have learned how to use optical scanners and computers to aggregate and analyze test scores. Only a very small fraction of the time and talent that has been devoted to the development and perfection of written tests has been applied in efforts to improve any of the possible alternatives to testing.

No one who attends to news of educational issues can be unaware of the attacks that Ralph Nader, the National Educational Association, and various ad hoc groups have been making on educational testing recently. Seeing so much smoke, the public might conclude that something is seriously wrong. But artfully managed, a very small fire can produce a great deal of smoke. Most of the criticisms of educational tests are ill founded, based on inadequate information or deliberate misrepresentation. They are certain to be recognized sooner or later as specious and unsound.

Those who make and use tests for selection and certification are no more secretive than is required to maintain test validity and economy of operation. Test score differences among racial and socio-economic groups in society are not evidence of test bias. They simply report real differences in educational development among these groups. Scores on the

tests can be raised somewhat by coaching, but not nearly enough to make a poor student look average in ability or an average student look good.

The tests measure achievement in learning accurately. A good standardized test taken by a typical group of students can differentiate thirty levels of ability consistently enough so that if the examinees were retested with an alternate form of the test, 50% of them would be placed in the same level or just one level above or below. For only one examinee in ten would the difference be more than three of the thirty levels. Predictions of future success are likely to be less accurate than this because the ability a person has developed through learning, while it is a vitally important factor, is only one of the factors that contribute to future success.

Thus despite the criticisms that have been widely publicized, written tests can measure the extent to which pupils have learned. The teacher is responsible for creating and maintaining a classroom environment that will maximize learning. Tests can reveal how effectively the teacher has discharged this basic responsibility. Some of the tests a teacher ought to use will be external standardized tests. Such tests reflect a concensus of good teachers on what all pupils ought to be learning. They also permit comparisons of the achievements of one teacher's pupils with the achievements of other similar pupils. Other tests a teacher ought to use will be designed and built by the teacher. Any good teacher will have some special things to teach that are not tested adequately or at all on a particular standardized test. To present a complete assessment of effectiveness, a teacher will need to

use not just standardized tests but teacher-made tests as well. That many college students are given certificates to teach without having had any systematic training in how to make or use tests is scandalous.

The tasks included in the standardized test a teacher selects, and in the classroom tests the teacher builds, indicate as clearly as can be indicated what has been set out to be learned. The success of the pupils in handling those tasks give evidence on how well they have learned it. Thus tests of achievement in learning serve several functions. They reveal to the school and to the community what the teacher was trying to teach. They reveal how well the pupils learned those things. They reveal how successfully the teacher taught.

This argument for the use of tests to measure how much pupils have learned rests on the premise that what is to be learned is mainly useful verbal knowledge. Suppose that premise is rejected. Then the one who rejects it must not only identify what other acquisition is more desirable. He must also describe how amounts of it can be measured. For an objective whose degree of attainment can not be assessed is useless. If reaching a goal is all the same as not reaching it, if coming closer to it is all the same as staying farther away from it, what is the point of trying to reach it or approach it? But if, on the other hand, reaching the goal or coming closer to it does make a difference, then the degree of attainment can be measured. For all that measurement requires, basically, is the observation of differences, the ability to distinguish more from less. Thus any alternative to knowledge as a goal of instruction must not only be identified, it must also be shown to be measurable. No such

alternative has yet been identifed.

4. Who should do the measuring?

Who should collect and present evidence on what was set out to be learned and how well it was learned? There are several reasons for believing that the teachers themselves should be responsible for providing evidence of their effectiveness. Accepting this responsibility is more consistent with the teacher's status as a professional, of the teacher's role as a largely independent manager of the instructional process. Teachers are not, and should not be, subject to detailed direction of their classroom activities.

If the teachers themselves have the opportunity to choose what evidence of effectiveness to provide, and how to collect it and to present it, they will have more flexibility in fitting the assessment to the instruction. They will have better opportunities to point out special strengths and to explain apparent weaknesses. Finally, being themselves in charge of the assessment process, they are likely to find it much more acceptable.

Of course, the assessment data will need to be presented to and reviewed by the instructional supervisor. This presentation will provide the occasion for the periodic reviews of a teacher's performance that are essential to the management of an effective educational institution. Among the questions that might be discussed are the following:

How adequate is the evidence of pupil learning?

How appropriate are the teacher's instructional goals?

How effective are the teacher's instructional methods?

What could the school do to help the teacher maintain a better environment for learning in the classroom?

Too often, in contemporary schools, the administrators are so busy administering that they neglect their important responsibilities as supervisors of instruction. Meetings with each teacher to review the teacher's assessment of pupil learnings would provide the occasions and the incentives for instructional supervision. The administrator may know no more about good teaching then the teacher he is talking to, but he does have authority and resources. The discussion of instructional goals, outcomes and problems with him is almost certain to be mutually beneficial. It is also likely to move the school along the road to educational excellence.

In conclusion, let me ask you to recall the proposition with which this paper began: almost all teachers should be held responsible for providing evidence on how much their pupils have learned of what was set out to be learned. Further, let me ask you to recall the main arguments in support of this proposition, namely, (1) that pupils should be concerned mainly with developing structures of useful verbal knowledge; (2) that pupil learning ought to be measured to motivate and reward pupil efforts to learn and to motivate and reward the efforts of their teachers to facilitate learning; (3) that written tests afford the best available means for measuring pupil achievements in learning; and (4) that teachers should select or construct the tests to be used, administer them, and report and interpret the results. There is no insurmountable obstacle to doing any of these things, and there are no alternative goals or means

of education that would be likely to attract and hold more public support for our schools. If we seek excellence in learning, there are no better policies to adopt.